Neuronal enhancers are hotspots for DNA single-strand break repair

Article (Accepted Version)

nature

https://doi.org/10.1038/s41586-021-03468-5

Accelerated Article Preview

Neuronal enhancers are hotspots for DNA single-strand break repair

Received: 28 November 2020

Accepted: 17 March 2021

Accelerated Article Preview Published online 25 March 2021

Cite this article as: Wu, W. et al. Neuronal enhancers are hotspots for DNA single-strand break repair. *Nature* https://doi.org/10.1038/s41586-021-03468-5 (2021).

Wei Wu, Sarah E. Hill, William J. Nathan, Jacob Paiano, Elsa Callen, Dongpeng Wang, Kenta Shinoda, Niek van Wietmarschen, Jennifer M. Colón-Mercado, Dali Zong, Raffaella De Pace, Han-Yu Shih, Steve Coon, Maia Parsadanian, Raphael Pavani, Hana Hanzlikova, Solji Park, Seol Kyoung Jung, Peter J. McHugh, Andres Canela, Chongyi Chen, Rafael Casellas, Keith W. Caldecott, Michael E. Ward & André Nussenzweig

Article

Neuronal enhancers are hotspots for DNA single-strand break repair

https://doi.org/10.1038/s41586-021-03468-5

Received: 28 November 2020

Accepted: 17 March 2021

Published online: 25 March 2021

Wei Wu[1,12], Sarah E. Hill[2,12], William J. Nathan[1,3,12], Jacob Paiano[1], Elsa Callen[1], Dongpeng Wang[1], Kenta Shinoda[1], Niek van Wietmarschen[1], Jennifer M. Colón-Mercado[2], Dali Zong[1], Raffaella De Pace[4], Han-Yu Shih[5], Steve Coon[4], Maia Parsadanian[2], Raphael Pavani[1], Hana Hanzlikova[6,7], Solji Park[8,9], Seol Kyoung Jung[8,9], Peter J. McHugh[3], Andres Canela[10], Chongyi Chen[11], Rafael Casellas[8,9], Keith W. Caldecott[6,7]✉, Michael E. Ward[2]✉ & André Nussenzweig[1]✉

Defects in DNA repair frequently lead to neurodevelopmental and neurodegenerative diseases, underscoring the particular importance of DNA repair in long-lived post-mitotic neurons[1,2]. The cellular genome is subjected to a constant barrage of endogenous DNA damage, but surprisingly little is known about the identity of the lesion(s) that accumulate in neurons and whether they accrue throughout the genome or at specific loci. Here we show that post-mitotic neurons accumulate unexpectedly high levels of DNA single-strand breaks (SSBs) at specific sites within the genome. Genome-wide mapping reveals that SSBs are located within enhancers at or near CpG dinucleotides and sites of DNA demethylation. These SSBs are repaired by PARP1 and XRCC1-dependent mechanisms. Notably, deficiencies in XRCC1-dependent short-patch repair increase DNA repair synthesis at neuronal enhancers, whereas defects in long-patch repair reduce synthesis. The high levels of SSB repair in neuronal enhancers are therefore likely to be sustained by both short-patch and long-patch processes. These data provide the first evidence of site- and cell type-specific SSB repair, revealing unexpected levels of localized and continuous DNA breakage in neurons. In addition, they suggest an explanation for the neurodegenerative phenotypes that occur in patients with defective SSB repair.

Recurrant sites of DNA repair in neurons

An obligatory and characteristic step of DNA repair is gap filling, in which excised or missing nucleotides are replaced using the undamaged strand as a template[3]. If a sufficient number of nucleotides are incorporated, DNA repair synthesis can be used as a proxy of the extent and location of endogenous DNA damage[4]. In light of this, we developed a method to map sites of DNA repair synthesis by sequencing (synthesis-associated with repair sequencing; SAR-seq). We labeled post-mitotic iPSC-derived glutamatergic neurons (i^{3}Neurons[5,6]) on day 6 post-differentiation with EdU for 18 hours, biotinylated the labeled DNA, reduced it to 150-200 bp by sonication, and then isolated the biotinylated DNA for high-throughput sequencing (Extended Data Fig. 1a). We identified >55,000 SAR-seq peaks at recurrent genomic locations in neurons, which were highly reproducible between different experiments (Fig. 1a, Extended Data Fig. 1b-d). Peaks were not caused by DNA synthesis during S phase[7,8] because i^{3}Neurons are post-mitotic (Extended Data Fig. 1e) and the SAR-seq peaks were unaffected by inhibition of the replicative DNA polymerase α (Extended Data Fig. 2a-c). In contrast and as expected, neuronal SAR was largely prevented by hydroxyurea (HU), which reduces the availability of deoxyribonucleotides (Extended Data Fig. 2 a, b).

The SAR-seq peaks were ~200-2,000 bp in width (averaging 901 bp) (Extended Data Fig. 2d), with each peak presumably comprising multiple clustered sites of DNA repair (see below). The most prominent neuronal SAR-seq peaks were detectable by pulse labeling with EdU for just 1 hour, and EdU incorporation approached saturation after labeling for 18 hours (Extended Data Fig. 2e, f).

When iPSCs were differentiated into skeletal muscle cells, we did not detect incorporation of EdU despite labeling cells for 18 hours (Extended Data Fig. 3a). Similarly, we failed to detect EdU incorporation in G0-arrested pre-B cells, although we could detect EdU incorporation in pre-B cells after inducing site-specific DNA double strand breaks (DSBs) (Extended Data Fig. 3b). To rule out the possibility that

[1]Laboratory of Genome Integrity, National Cancer Institute, NIH, Bethesda, MD, USA. [2]National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD, USA. [3]Department of Oncology, MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, United Kingdom. [4]Eunice Kennedy Shriver National Institute of Child Health and Human Development, Rockville, MD, USA. [5]National Eye Institute, NIH, Bethesda, MD, USA. [6]Department of Genome Dynamics, Institute of Molecular Genetics of the Czech Academy of Sciences, 142 20, Prague, 4, Czech Republic. [7]Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Brighton, BN1 9RQ, UK. [8]Lymphocyte Nuclear Biology, National Institute of Arthritis and Musculoskeletal and Skin Diseases and National Cancer Institute, NIH, Bethesda, MD, USA. [9]NIH Regulome Project, NIH, Bethesda, MD, USA. [10]The Hakubi Center for Advanced Research and Radiation Biology Center, Graduate School of Biostudies, Kyoto University, Kyoto, Japan. [11]Laboratory of Biochemistry and Molecular Biology, National Cancer Institute, NIH, Bethesda, MD, USA. [12]These authors contributed equally: Wei Wu, Sarah E. Hill, William J. Nathan. ✉e-mail: k.w.caldecott@sussex.ac.uk; wardme@nih.gov; andre_nussenzweig@nih.gov

the SAR-seq peaks were an artifact of iPSC differentiation, we labelled *bona fide* rat neurons with EdU. Similar to i^{3}Neurons, we detected robust peaks of EdU incorporation at 22,196 specific sites in rat neurons (Fig. 1b, Extended Data Fig. 5e). Thus, the high frequency of recurrent DNA synthesis appears to be a specific feature of post-mitotic neurons.

DNA repair sites enriched at enhancers

Neuronal SAR-seq peaks were enriched in intragenic regions (Extended Data Fig. 3c, d) and further within expressed genes (Extended Data Fig. 3d, e). However, SAR signal intensity did not correlate with transcript levels (Extended Data Fig. 3e). Moreover, the sites of EdU incorporation were not associated with strand-specificity, as EdU was incorporated uniformly in both transcribed and non-transcribed strands (Extended Data Fig. 3f).

We next searched for specific DNA motifs among the strongest 5,000 SAR-seq peaks. More than 25% of the sites harbored a motif similar to the ONECUT family of transcription factors, which was centered at SAR-seq peak summits (Extended Data Fig. 4a). Since ONECUT1 can promote genomic accessibility in neurons[9], we compared SAR-seq peaks with accessible regions using ATAC-seq. 54% of all SAR-seq regions coincided with ATAC-seq peaks (Fig. 1a, Extended Data Fig. 4b, c), and the widths of SAR-seq and ATAC-seq peaks were correlated (Extended Data Fig. 4d), suggesting that open chromatin structure influences the extent of DNA synthesis.

Despite their localization in open chromatin, SAR-seq peaks were not enriched at promoters (Extended Data Fig. 4e-g), which exhibited only modest levels of DNA synthesis. However, we detected a strong correlation between locations of DNA synthesis and of neuronal enhancers, as measured by ChIP-seq for H3K4me1, H3K27ac and MLL4 (Fig. 1a, Extended Data Fig. 5a, b). Nevertheless, SAR-seq peaks were not strongly correlated with other, non-enhancer-specific epigenetic indicators of open chromatin or condensed chromatin (Extended Data Fig. 5a, c). Similar to i^{3}Neurons, SAR-seq peaks in primary rat neurons overlapped and correlated with rat H3K4me1 ChIP-seq (Extended Data Fig. 5 d-f).

We performed ultra-deep Hi-C in i^{3}Neurons to examine chromosomal interactions associated with SAR-seq peaks within the context of topological associated domains (TADs) (Extended Data Fig. 6a). Enhancers with SAR-seq peaks had more intra-TAD interactions than enhancers without SAR-seq peaks. Consistently, H3K27ac-a mark of active enhancers-was significantly enriched at enhancers with SAR-seq peaks (Extended Data Fig. 6a). Chromatin looping interactions detected by promoter capture Hi-C (pcHi-C) enables the linkage of distal enhancer elements to their target genes[10]. Compared to random sites, we observed an 8-fold enrichment of SAR sites among *in vivo* validated enhancer elements[11] that overlapped with neuronal H3K4me1 histone marks and a 4-fold enrichment using a pcHi-C dataset generated in i^{3}Neurons (Extended Data Fig 6b)[10]. The latter includes enhancers whose activity was validated in human neurons using CRISPR techniques (Extended Data Fig 6c, d). Thus, distal-acting regulatory enhancers are hotspots of neuronal SAR.

To determine whether the enhancers with SAR-seq peaks are specific to neurons, we compared sites of H3K4me1 in i^{3}Neurons and iPSCs. Only 2% of the SAR-seq peaks overlapped with iPSC-specific H3K4me1 peaks, while most of the SAR-seq peaks overlapped with either neuron-specific or shared H3K4me1 sites (Extended Data Fig. 6e). Thus, SAR is associated with enhancers that are active in differentiated neurons.

Gene Ontology (GO) analysis of the genes containing SAR-seq peaks revealed than an enrichment of GO terms related to nervous system function (Extended Data Fig. 6f). We then further characterized the neuronal enhancers that were associated with SAR. Super-enhancers (SE) are a large collection of enhancers that drive transcription of genes involved in cell identity. Using H3K27ac ChIP-seq, we found approximately 1,300 SE in i^{3}Neurons. Interestingly, 90% of SE exhibited SAR-seq peaks, whereas less than 25% of conventional enhancers possessed SAR-seq peaks (Extended Data Fig. 6g, h). Collectively, these data identify enhancers and genes associated with neuronal function as hotspots of recurrent DNA synthesis.

PARP activity at neuronal enhancers

Given the close association between unrepaired DNA strand breaks and neurodegeneration[1,2], we wondered if the sites of EdU incorporation might reflect sites of DNA break repair. To test this, we measured the activity of poly(ADP-ribose) polymerases (PARPs) at the sites of DNA synthesis. Various types of DNA damage including SSBs, DSBs, and single-strand gaps activate PARP1 and PARP2[12,13]. PARP activity signals the presence of these lesions by modifying localized proteins with poly(ADP-ribose) (PAR)[12,13]. When we monitored ADP-ribosylation in individual neurons[14], we detected focal sites of nuclear ADP-ribose. These focal sites, along with pan-nuclear staining, increased following treatment with the genotoxin methylmethansulfate (MMS) (Fig. 1c, Extended Data Fig. 7a). To determine whether the sites of endogenous ADP-ribosylation were localized to sites of neuronal DNA synthesis, we employed ADP-ribose ChIP-seq. As a positive control, we first confirmed could detect ADP-ribosylation occurring at site-specific DSBs (Extended Data Fig. 7b). In i^{3}Neurons, we found that the endogenous sites of ADP-ribosylation co-localized with SAR-seq peaks (Fig. 1d, Extended Data Fig. 7c), suggesting that the recurrent sites of DNA synthesis are associated with DNA strand break repair.

DNA repair is not associated with DSBs

Neuronal activity has been reported to cause DSBs generated by topoisomerase 2 (TOP2)[15,16]. TOP2-induced DSBs can promote the expression of early response genes[15], associating these DNA breaks with regions of transcriptional activity[17,18]. Treatment of i^{3}Neurons with etoposide to trigger TOP2-induced DSBs resulted in DNA synthesis within gene bodies (Extended Data Fig. 8a, b). However, most of the sites of etoposide-induced DNA synthesis were distinct from those detected in untreated neurons (Extended Data Fig. 8a, b). In addition, we did not detect DSBs in unchallenged i^{3}Neurons as measured by either γ-H2AX/53BP1 immunostaining (Extended Data Fig. 8c) or by END-seq (Extended Data Fig. 8d)[19]. Thus, sites of DNA synthesis in neuronal enhancers are independent of DSBs.

SSBs at sites of DNA synthesis

In addition to DSBs, PARP1 and/or PARP2 are also activated at SSBs and subsequently recruit the XRCC1 protein complex that accelerates SSB repair[20,21]. We therefore examined the genomic localization of XRCC1 by ChIP-seq. Similar to sites of PARP activity, XRCC1 co-localized with SAR-peaks, and the intensity of XRCC1 binding correlated with the intensity of EdU incorporation in both human i^{3}Neurons and rat primary neurons (Fig. 1d, Extended Data Fig. 7c and Extended Data 8e-g). Thus, sites of DNA synthesis colocalize with sites of PARP activation and XRCC1-associated SSB repair.

To directly map SSBs at nucleotide resolution, we treated agarose-embedded i^{3}Neuron plugs with recombinant single strand-specific S1 nuclease *in situ* to convert SSBs into DSBs, which we then detected using END-seq. To test whether this method would successfully convert DNA nicks into DSBs, we first incubated the i^{3}Neuron agarose plugs with the site-specific nicking endonuclease Nt.BspQI prior to S1 nuclease treatment. Nt.BspQI nicks were readily converted into DSBs by S1 (Extended Data Fig. 9a). Moreover, S1-END-seq could detect nicks generated by an inducible CAS9 nickase expressed in G1-arrested MCF10 cells (Extended Data Fig. 9b,c). However, S1-END-seq alone failed to detect endogenous SSBs in i^{3}Neurons (Fig 2a, Extended Data Fig. 9a).

We surmised that failure to detect endogenous SSBs in neurons could be due to their rapid repair. To increase the half-life of SSBs, we

incubated i^{3}Neurons with a mixture of chain terminating dideoxynucleosides (ddA+ddC+ddT+ddG, denoted ddN). Strikingly, this led to robust SSB accumulation as detected by S1-END-seq (Fig. 2a), which colocalized extensively with SAR-seq peaks (Fig. 2a, b, and Extended Data Fig. 9d, e). Many of the SAR-seq peaks contained multiple S1-END-seq peaks (Fig. 2a, **inset**). Each peak spanned an average of ~25 nucleotides (Extended Data Fig. 9f), thus comprising multiple-clustered single-strand gaps. Interestingly, we also detected a very low-level of DSBs at SAR-seq peaks following chain termination, which were independent of S1 nuclease treatment and likely reflected closely apposed SSBs on opposite DNA strands (Fig. 2b).

S1-END-seq enabled the identification of the location of the endogenous SSBs with much higher resolution than did SAR-seq, which is limited to the size of the sonicated fragments (150-200 nt). The improved resolution revealed a prevalence of C/G nucleotides at S1-END-seq peak summits on positive or negative strands (Fig 2c), and that CpG dinucleotides were highly enriched at SAR sites (Extended Data Fig. 9g, h). Collectively, these data provide direct evidence for the site-specific formation of clustered SSBs in neuronal enhancers at or near C/G nucleotides.

SAR involves long-patch SSB repair

SSB repair comprises both short-patch and long-patch sub pathways, in which single or multiple nucleotides are replaced at the site of the SSB, respectively[22,23]. PARP1 and XRCC1 promote the repair of a wide spectrum of SSBs, primarily by short-patch repair. We therefore examined the impact of inhibiting and/or depleting these proteins on neuronal DNA synthesis. We observed a reproducible increase in EdU incorporation at SAR sites if neurons were co-incubated with any of three independent inhibitors of PARP1, or if PARP1 was depleted using CRISPR interference (CRISPRi)[24] (Fig. 3a and Extended Data Fig. 10a,b). Depletion of XRCC1 similarly led to a prominent increase in EdU incorporation at sites of SAR (Fig. 3b and Extended Data Fig. 10c- e). These data suggest that if PARP1/XRCC1-dependent short-patch SSB repair is impeded, compensatory long-patch SSB repair is increased.

During short-patch SSB repair, a single nucleotide is replaced at the site of the break[22,23,25]. Nucleotide replacement typically requires DNA polymerase β (POLβ), which interacts directly with XRCC1[20]. In contrast, during long-patch repair, alternative DNA polymerases such as POLε and POLδ can generate a 2-20 nucleotide DNA repair patch. Consistent with this, we found that depletion of POLβ resulted in a dramatic increase in DNA synthesis at neuronal enhancers (Fig. 3c; Extended Data Fig. 10f-h). In contrast, incubation with aphidicolin (Aph), an inhibitor of POLε and POLδ, greatly reduced SAR in both wild-type and POLβ-depleted neurons when used at concentrations (50μM) that inhibit cellular repair synthesis[4,26] (Fig. 3c; Extended Data Fig. 10i). Collectively, these data indicate that neuronal site-specific SSBs are repaired by both short-patch and long-patch repair, with the latter pathway being the primary source of SAR-seq signal.

SSBs and active cytosine demethylation

The detection of DNA repair synthesis at neuronal enhancers could indicate that the latter are sites of increased DNA damage or, alternatively, that they are sites of preferential DNA repair. Our observations using S1-END-seq, in which we see that SSBs are enriched at neuronal enhancers, strongly supports the first hypothesis. To test this further, we treated i^{3}Neurons with MMS to introduce SSBs stochastically across the genome. We predicted that if DNA repair is targeted to enhancer sites, induction of random SSBs across the genome should not affect the location of the SAR-seq peaks. However, despite the expected increase in the overall level of EdU incorporation 2-5 fold, MMS ablated the appearance of SAR-seq peaks at enhancers. This observation suggests that elevated levels of SSBs occur specifically at neuronal enhancers and account for the peaks of DNA repair synthesis at these sites, rather than a selective and site-specific repair process (Extended Data Fig. 11a).

What is the source of the localized SSBs? Oxidative DNA lesions such as 8-oxoguanine are strongly implicated sources of DNA damage in the brain, but unlike the sites of DNA repair synthesis detected here such lesions have not been shown to preferentially accumulate in open chromatin or at enhancers[27]. In contrast, enhancers may be especially vulnerable to SSBs induced by TOP1, a topoisomerase enzyme that is implicated in enhancer activation[28,29]. However, such SSBs are also unlikely to be the source of the localized SSBs, since depletion of the polynucleotide kinase 3′-phosphatase (PNKP)-which is recruited by XRCC1 to repair TOP1-induced SSBs[30,31]- had only a small impact on the SAR-seq peaks (Extended Data Fig. 11 b, c). This was in contrast to the DNA repair synthesis of *bona fide* TOP1-induced SSBs that we triggered by treating neurons with the TOP1 poison camptothecin, which were increased by PNKP depletion and located in gene bodies rather than in enhancers (Extended Data Fig. 11d).

Given our finding that SSBs associated with SAR-seq peaks were enriched at C/G nucleotides (Fig. 2c and Extended Data Fig. 9g, h), we entertained the possibility that these SSBs were sites of cytosine demethylation (Fig. 4c). Consistent with this idea, active DNA demethylation of cytosine at CpG sites occurs preferentially at enhancers[32], is 10-fold more active in post-mitotic neurons than peripheral cell types[33], and generates SSBs that are intermediates of XRCC1-associated, but not PNKP-associated, BER[32,34]. Active demethylation via ten-eleven translocation (TET) family (TET1, TET2, and TET3) enzymes is initiated through progressive oxidation of 5mC to 5hmC, 5fC, or 5caC, and steady state levels of 5hmC account for approximately 40% of modified cytosines in the brain[33]. Using labeling methods to detect 5hmC and 5fC[35,36], we mapped oxidized forms of 5-methylcytosine genome-wide in i^{3}Neurons. We found that the sites of both DNA repair synthesis (measured by SAR-seq) and SSBs (measured by S1-END-seq) overlapped with peaks of 5hmC and 5fC, and the intensity of SSBs correlated with that of 5hmC and 5fC (Fig. 4a, b and Extended Data Fig. 11e, f). While future studies will be needed to determine the precise source(s) of SSBs at neuronal enhancers, our data implicate cycles of cytosine methylation and demethylation at neuronal enhancers as a potential source of neuronal site-specific DNA single-strand breakage.

Conclusions

Our study reveals that human post-mitotic neurons are subject to an unexpected level of localized DNA synthesis that is associated with ongoing sites of SSB repair at neuronal enhancers. The scaffold protein XRCC1 is of particular importance during SSB repair because it is physically associated with a variety of SSB repair enzymes including TDP1, POLβ, PNKP, APTX, and DNA ligase III (LIG3)[20], some of which if mutated result in hereditary neurodegenerative diseases[1,2]. Failure to repair the site-specific SSBs that we have discovered here may thus contribute to such diseases, though we do not rule out an involvement of other SSB sources. Unrepaired SSBs may trigger neuropathology in several ways, such as via excessive or aberrant activation of PARP1[37]. In addition, our current data raise the possibility that an increased dependency on DNA repair synthesis at specific sites of the genome increases the mutational burden in long-lived neurons[38]. For example, it is possible that an accrual of mutations at sites of recurrent DNA repair synthesis within enhancers could lead to aberrant gene expression, resulting in neurological dysfunction over time.

In summary, we describe new methods that enable genome-wide mapping of endogenous sites of frequent DNA single-strand breakage and repair synthesis in post-mitotic neurons. Our findings identify enhancers as hotspots of DNA damage in human post-mitotic neurons, perhaps explaining why DNA SSB repair is important for neurological functionality during development and normal ageing.

Note: During the preparation of this manuscript, we became aware of the closely related work of D. Reid *et al.*, which demonstrated recurrent DNA repair sites in embryonic stem cell-derived neurons[39].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03468-5.

1. Caldecott, K. W. Single-strand break repair and genetic disease. *Nat Rev Genet* **9**, 619-631, https://doi.org/10.1038/nrg2380 (2008).
2. McKinnon, P. J. Genome integrity and disease prevention in the nervous system. *Genes Dev* **31**, 1180-1194, https://doi.org/10.1101/gad.301325.117 (2017).
3. Tubbs, A. & Nussenzweig, A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* **168**, 644-656, https://doi.org/10.1016/j.cell.2017.01.002 (2017).
4. Miller, M. R. & Chinault, D. N. The roles of DNA polymerases alpha, beta, and gamma in DNA repair synthesis induced in hamster and human cells by different DNA damaging agents. *J Biol Chem* **257**, 10204-10209 (1982).
5. Fernandopulle, M. S. *et al*. Transcription Factor-Mediated Differentiation of Human iPSCs into Neurons. *Curr Protoc Cell Biol* **79**, e51, https://doi.org/10.1002/cpcb.51 (2018).
6. Wang, C. *et al*. Scalable Production of iPSC-Derived Human Neurons to Identify Tau-Lowering Compounds by High-Content Screening. *Stem Cell Reports* **9**, 1221-1233, https://doi.org/10.1016/j.stemcr.2017.08.019 (2017).
7. Macheret, M. & Halazonetis, T. D. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature* **555**, 112-116, https://doi.org/10.1038/nature25507 (2018).
8. Tubbs, A. *et al*. Dual Roles of Poly(dA:dT) Tracts in Replication Initiation and Fork Collapse. *Cell* **174**, 1127-1142 e1119, https://doi.org/10.1016/j.cell.2018.07.011 (2018).
9. van der Raadt, J., van Gestel, S. H. C., Nadif Kasri, N. & Albers, C. A. ONECUT transcription factors induce neuronal characteristics and remodel chromatin accessibility. *Nucleic Acids Res* **47**, 5587-5602, https://doi.org/10.1093/nar/gkz273 (2019).
10. Song, M. *et al*. Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat Genet* **51**, 1252-1262, https://doi.org/10.1038/s41588-019-0472-1 (2019).
11. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser–a database of tissue-specific human enhancers. *Nucleic Acids Res* **35**, D88-92, https://doi.org/10.1093/nar/gkl822 (2007).
12. Gupte, R., Liu, Z. & Kraus, W. L. PARPs and ADP-ribosylation: recent advances linking molecular functions to biological outcomes. *Genes Dev* **31**, 101-126, https://doi.org/10.1101/gad.291518.116 (2017).
13. Hanzlikova, H. & Caldecott, K. W. Perspectives on PARPs in S Phase. *Trends Genet* **35**, 412-422, https://doi.org/10.1016/j.tig.2019.03.008 (2019).
14. Gibson, B. A., Conrad, L. B., Huang, D. & Kraus, W. L. Generation and Characterization of Recombinant Antibody-like ADP-Ribose Binding Proteins. *Biochemistry* **56**, 6305-6316, https://doi.org/10.1021/acs.biochem.7b00670 (2017).
15. Madabhushi, R. *et al*. Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. *Cell* **161**, 1592-1605, https://doi.org/10.1016/j.cell.2015.05.032 (2015).
16. Suberbielle, E. *et al*. Physiologic brain activity causes DNA double-strand breaks in neurons, with exacerbation by amyloid-beta. *Nat Neurosci* **16**, 613-621, https://doi.org/10.1038/nn.3356 (2013).
17. Canela, A. *et al*. Topoisomerase II-Induced Chromosome Breakage and Translocation Is Determined by Chromosome Architecture and Transcriptional Activity. *Mol Cell* **75**, 252-266 e258, https://doi.org/10.1016/j.molcel.2019.04.030 (2019).
18. Gomez-Herreros, F. *et al*. TDP2 suppresses chromosomal translocations induced by DNA topoisomerase II during gene transcription. *Nat Commun* **8**, 233, https://doi.org/10.1038/s41467-017-00307-y (2017).
19. Canela, A. *et al*. DNA Breaks and End Resection Measured Genome-wide by End Sequencing. *Mol Cell* **63**, 898-911, https://doi.org/10.1016/j.molcel.2016.06.034 (2016).
20. Caldecott, K. W. XRCC1 protein; Form and function. *DNA Repair (Amst)* **81**, 102664, https://doi.org/10.1016/j.dnarep.2019.102664 (2019).
21. Hanzlikova, H., Gittens, W., Krejcikova, K., Zeng, Z. & Caldecott, K. W. Overlapping roles for PARP1 and PARP2 in the recruitment of endogenous XRCC1 and PNKP into oxidized chromatin. *Nucleic Acids Res* **45**, 2546-2557, https://doi.org/10.1093/nar/gkw1246 (2017).
22. Caldecott, K. W. DNA single-strand break repair. *Exp Cell Res* **329**, 2-8, https://doi.org/10.1016/j.yexcr.2014.08.027 (2014).
23. Caldecott, K. W. Mammalian DNA base excision repair: Dancing in the moonlight. *DNA Repair (Amst)* **93**, 102921, https://doi.org/10.1016/j.dnarep.2020.102921 (2020).
24. Tian, R. *et al*. CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron* **104**, 239-255 e212, https://doi.org/10.1016/j.neuron.2019.07.014 (2019).
25. Beard, W. A., Horton, J. K., Prasad, R. & Wilson, S. H. Eukaryotic Base Excision Repair: New Approaches Shine Light on Mechanism. *Annu Rev Biochem* **88**, 137-162, https://doi.org/10.1146/annurev-biochem-013118-111315 (2019).
26. DiGiuseppe, J. A., Hunting, D. J. & Dresler, S. L. Aphidicolin-sensitive DNA repair synthesis in human fibroblasts damaged with bleomycin is distinct from UV-induced repair. *Carcinogenesis* **11**, 1021-1026, https://doi.org/10.1093/carcin/11.6.1021 (1990).
27. Poetsch, A. R. The genomics of oxidative DNA damage, repair, and resulting mutagenesis. *Comput Struct Biotechnol J* **18**, 207-219, https://doi.org/10.1016/j.csbj.2019.12.013 (2020).
28. Bansal, K., Yoshida, H., Benoist, C. & Mathis, D. The transcriptional regulator Aire binds to and activates super-enhancers. *Nat Immunol* **18**, 263-273, https://doi.org/10.1038/ni.3675 (2017).
29. Puc, J. *et al*. Ligand-dependent enhancer activation regulated by topoisomerase-I activity. *Cell* **160**, 367-380, https://doi.org/10.1016/j.cell.2014.12.023 (2015).
30. Kalasova, I. *et al*. Pathological mutations in PNKP trigger defects in DNA single-strand break repair but not DNA double-strand break repair. *Nucleic Acids Res* **48**, 6672-6684, https://doi.org/10.1093/nar/gkaa489 (2020).
31. Whitehouse, C. J. *et al*. XRCC1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair. *Cell* **104**, 107-117, https://doi.org/10.1016/s0092-8674(01)00195-7 (2001).
32. Lio, C. J. *et al*. TET methylcytosine oxidases: new insights from a decade of research. *J Biosci* **45** (2020).
33. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929-930, https://doi.org/10.1126/science.1169786 (2009).
34. Steinacher, R. *et al*. SUMOylation coordinates BERosome assembly in active DNA demethylation during cell differentiation. *EMBO J* **38**, https://doi.org/10.15252/embj.201899242 (2019).
35. Song, C. X. *et al*. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678-691, https://doi.org/10.1016/j.cell.2013.04.001 (2013).
36. Szulwach, K. E. *et al*. 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat Neurosci* **14**, 1607-1616, https://doi.org/10.1038/nn.2959 (2011).
37. Hoch, N. C. *et al*. XRCC1 mutation is associated with PARP1 hyperactivation and cerebellar ataxia. *Nature* **541**, 87-91, https://doi.org/10.1038/nature20790 (2017).
38. Lodato, M. A. *et al*. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555-559, https://doi.org/10.1126/science.aao4426 (2018).
39. Ried, D. A. *et al*. Incorporation of a nucleoside analog maps genome repair sites in post-mitotic human neurons. *Science* (2021), *in press*.

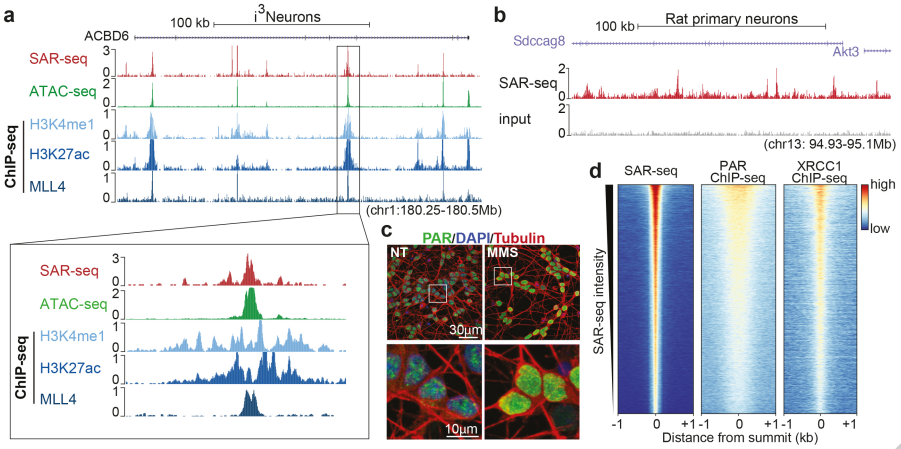


Fig. 1 | SAR-seq peaks occur within enhancers and are associated with PARP activation. a) Genome browser screenshots of SAR-seq (n=3), ATAC-seq (n=1), H3K4me1 (n=2), H3K27ac (n=1), and MLL4 (n=1) ChIP-seqs in i^{3}Neurons. *Inset*, zoomed-in screenshot of the indicated region highlighting overlapping peaks. **b)** Genome browser screenshot of SAR-seq performed in rat primary neurons (n=1) as well as input. The culture was co-incubated with 5μM aphidicolin to block DNA replication of S phase glial cells. **c)** Representative images of anti-poly(ADP-ribose) immunofluorescence staining (PAR, green) and neuronal marker Tubulin Beta 3 (red) in i^{3}Neurons counterstained with DAPI (blue) (data are representative of three independent experiments). As a positive control, cells were treated with 0.1 mg/ml MMS for 15 min. Boxed regions in top panels are enlarged in bottom panels. **d)** Heatmaps of SAR-seq, XRCC1 and PAR ChIP-seq signal ±1kb surrounding SAR-seq peak summits in i^{3}Neurons, ordered by SAR-seq intensity.

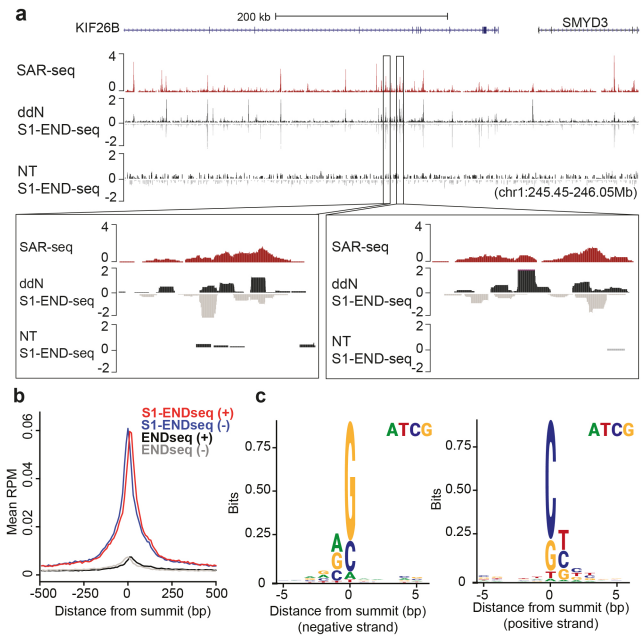


Fig. 2 | Single-strand breaks detected by S1-END-seq following ddN incubation. a) Genome browser screenshots illustrating the overlap in i^{3}Neurons between sites of SAR-seq and SSBs (S1-END-seq) detected in the presence of ddN (n=1) to block DNA ligation, or in its absence ("NT", n=1). S1-END-seq signal was separated by positive (black) and negative (grey) strands. *Inset*, zoomed screenshots demonstrating the presence of multiple SSBs within one SAR-seq peak. **b)** Aggregate plots of S1-END-seq (red: positive-strand, blue: negative-strand) and END-seq in the presence of ddN (n=1) (black: positive-strand, grey: negative-strand) signal ±500 bp around SAR-seq peak summits. **c)** Composite DNA sequence motif within ±5 bp surrounding the SSB summits on the positive strand (right panel) and on the negative strand (left panel) of the 10,000 most prominent S1-END-seq peaks.

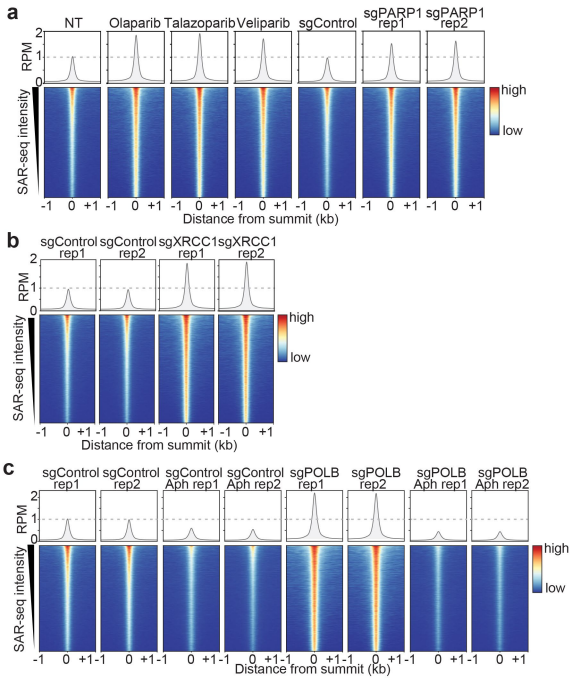


Fig. 3 | Localized SSB repair in neurons is comprised of short-patch and long-patch sub-pathways. a) Heatmaps of SAR-seq intensities ±1 kb surrounding SAR-seq peak summits for i^{3}Neurons treated with the indicated PARP inhibitors (n=3) or expressing non-targeted control (sgControl, n=1) or PARP1-targeted (sgPARP1, n=2) CRISPRi. Aggregate plots of SAR-seq intensity are shown in the top panel. NT: non-treated; rep: replicate. RPM: reads per million. **b)** Heatmaps of SAR-seq intensities ±1 kb surrounding SAR-seq peak summits for i^{3}Neurons expressing non-targeted control (sgControl, n=2) or XRCC1-targeted (sgXRCC1, n=2) CRISPRi. **c)** Heatmaps of SAR-seq intensities ±1kb surrounding SAR-seq peak summits for i^{3}Neurons expressing non-targeted (sgControl, n=2) or POLB-targeted (sgPOLB, n=2) CRISPRi. Where indicated, neurons were treated with 50 μM Aphidicolin (Aph) for 24 hours prior to and during EdU incorporation.

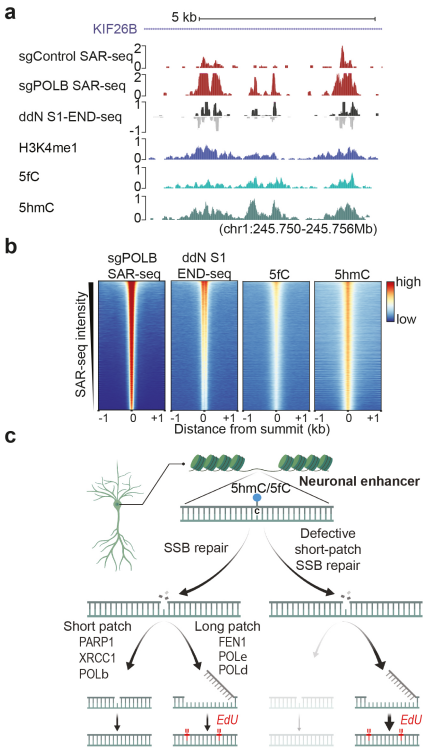


Fig. 4 | Localized SSB repair in neurons correlates with sites of oxidized 5-methylcytosine. a) Genome browser screenshot illustrating the overlap in i^{3}Neurons between sites of 5hmC (5hmC-SEAL, n=2), 5fC (5fC-SEAL, n=2), SSBs (ddN S1-END-seq), DNA repair synthesis (SAR-seq), and H3K4me1 ChIP-seq. Both positive (black) and negative (grey) S1-END-seq signal are shown. **b)** Heatmaps of signals for DNA repair synthesis (SAR-seq), SSBs (ddN S1-END-seq), 5fC (5fC-SEAL), and 5hmC (5hmC-SEAL) in i^{3}Neurons, plotted ±1 kb surrounding the summits of the SAR-seq peaks and ordered by SAR-seq intensity. **c)** Model depicting the balance between short-patch and long-patch SSB repair at neuronal enhancers. Note that while our data implicate SSBs arising during the base excision repair of modified cytosine residues within enhancers, we do not exclude the occurrence of other sites and sources of SSBs.

Methods

iPSC cell culture

All induced pluripotent stem cell (iPSC) experiments used the WTC11 line, which was derived from a healthy human male participant and obtained from the Coriell cell repository. All policies of the NIH Intramural Research Program for the registration and use of this iPSC line were followed. The WTC11 iPSC line was validated to have a normal male karyotype, as expected, and was confirmed to be mycoplasma free based on the Lonza "MycoAlert" mycoplasma testing kit. iPSC culture was performed as described previously[5]. Tissue culture treated dishes were coated with hESC-qualified matrigel (Corning, REF 354277). Matrigel was removed and iPSCs were plated in Essential 8 Medium (E8; Thermo Fisher Scientific, Cat. No. A1517001) and 10 µM ROCK inhibitor (RI; Y-27632; Selleckchem, Cat. No. S1049). iPSCs were maintained in a 37°C, 5% CO_2 incubator and fed every 1-2 days as needed. Cells were split using either accutase (Life Technologies, Cat. No. A1110501) for enzymatic dissociation into single cells or EDTA (0.5mM; Life Technologies, Cat. No. 15575020) for routine passaging. Media was supplemented with 10 µM RI to promote survival during passaging. As necessary iPSCs were frozen in 90% ES-qualified fetal bovine serum (FBS) (Sigma Aldrich, Cat. No. ES-009-B) and 10% DMSO (Mediatech Inc., Cat. No. 25-950-CQC), and then thawed rapidly at 37°C, followed by removal of FBS/DMSO and plating in E8+RI medium.

i^{3}Neuron culture

The human iPSCs used in this study were previously engineered[5,6] to express mouse Neurogenin-2 (NGN2) under a doxycycline-inducible promoter integrated at the AAVS1 safe harbor in the WTC11 background, +/- CAG-dCas9-BFP-KRAB at the CLYBL promoter[24]. For neuronal differentiation, 20-25 million iPSCs were plated on day 0 onto a 15 cm plate in N2 media composed of knockout DMEM/F12 media (Life Technologies Corporation, Cat. No. 12660012) with N2 supplement (Life Technologies Corporation, Cat. No. 17502048), 1x GlutaMAX (Thermofisher Scientific, Cat. No. 35050061), 1x MEM nonessential amino acids (NEAA) (Thermofisher Scientific, Cat. No. 11140050), 10 µM ROCK inhibitor (Y-27632; Selleckchem, Cat. No. S1049), and 2 µg/mL doxycycline (Clontech, Cat. No. 631311). N2 media was changed once a day for two more days. Day 3 cells were replated onto freshly-prepared poly-L-ornithine- (PLO; 0.1 mg/ml; Sigma, Cat. No. P3655-10MG) coated dishes as follows: Cells were washed with PBS, dissociated with accutase for 10 minutes at 37°C, washed and plated in i^{3}Neuron Culture Media: BrainPhys media (STEMCELL Technologies, Cat. No. 05790) supplemented with 1x B27 Plus Supplement (ThermoFisher Scientific, Cat. No. A3582801), 10 ng/mL BDNF (PeproTech, Cat. No. 450-02), 10 ng/mL NT-3 (PeproTech, Cat. No. 450-03), 1 mg/mL mouse laminin (Sigma, Cat. No. L2020-1MG), and 2 µg/mL doxycycline (Clontech, Cat. No. 631311). For 10 cm plates used in SAR-seq or CHIP-seq, 12-15 million neurons were plated. For 15 cm plates 30-45 million neurons were plated. For ibidi slides used in imaging experiments, 0.2 million neurons per well were plated. Unless otherwise noted, i^{3}Neurons were fed on day 6 during a half media change and harvested on day 7. For i^{3}Neurons cultured beyond 7 days, half media changes were conducted three times per week. In some experiments pre-differentiated i^{3}Neurons were frozen on day 3 in 90% fetal bovine serum (Sigma Aldrich, Cat. No. ES-009-B) and 10% DMSO (Mediatech Inc., Cat. No. 25-950-CQC), and then thawed rapidly at 37°C, followed by removal of FBS/DMSO and plating in i^{3}Neuron Culture Media. We did not detect any differences for experiments where day 3 neurons were thawed or plated immediately following differentiation.

iMuscle culture

To generate skeletal muscle myoblasts from human iPSCs, we engineered a doxycycline-inducible vector harboring a MyoD1 transcription factor transgene immediately followed by a co-inducible short hairpin RNA targeting Oct4 and selection marker/fluorescent protein cassette (MyoD-O iPSCs), similar to previous methods[40]. We used the PiggyBac™ system to facilitate delivery and genome integration of the transgene cassette. To increase transposase expression in iPSCs, we subcloned the PiggyBac™ transposase under a long version of the eF1a promoter, and co-transfected this transposase vector with the MyoD1 donor vector into iPSCs using lipofectamine stem (Thermofisher Scientific, Cat. No. STEM00015). Transfection rates were approximately

55% based on fluorescence of a co-expressed reporter gene, and we were able to obtain 100% MyoD1-O iPSCs after puromycin selection of iPSCs harboring stable integration of the transcription factor cassette.

To induce myogenic differentiation, MyoD-O iPSCs were dissociated using accutase (37°C for 10 minutes) and resuspended in myogenic induction media (MIM) and plated with 3x10^6 iPSCs per 10 cm dish. The MyoD-O iPSCs differentiation was modified from a recent report[41]. The MIM contained DMEM/F12 HEPES (Gibco, Cat. No. 11-330-032) supplemented to a final concentration of 1 mM sodium pyruvate, 1x MEM nonessential amino acids (NEAA) (Thermofisher Scientific, Cat. No. 11140050), 0.1 mM 2-mercaptoethanol (Gibco, Cat. No. 21985023), 10 µg/mL insulin (Roche, Cat. No. 11376497001), 2 µg/mL doxycycline and 10 µM RI.

After plating the MyoD-O iPSCs in MIM at day 0, fresh media was supplied daily until 80% confluency or more (day 6). Myoblast morphology was observed beginning at day 2 followed by myotube-like morphology at day 5. On day 6, the glycogen synthase kinase 3 (GSK-3) inhibitor, CHIR 99021, was administered for 48 hours to enhance myoblast differentiation and stimulate fusion (3 µM, Tocris, Cat. No. 4423) as reported by others[42,43]. At day 9 post-doxycycline, the differentiating myoblasts were harvested.

Lenti-virus production for CRISPRi guide delivery

Lenti-X Human Embryonic Kidney (HEK) cells were plated onto dishes coated with poly-L-ornithine (0.1 mg/ml PLO) at a density of 10-15 million cells (10 cm dish) or 30-45 million cells (15 cm dish) per dish in warm DMEM, high glucose GlutaMAX™ Supplement media (Life Technologies, Cat. No. 10566024) with 10% FBS (Sigma, Cat. No. TMS-013-B) and then cultured overnight to achieve approximately 90% confluency. The next morning cells were transfected using Lipofectamine 3000 (Life Technologies, Cat. No. L3000150). For each transfection, 2.4 mL of room temperature Opti-MEM media (Fisher Scientific, Cat. No. 31985062) and 60 µL of Lipo 3K were combined and incubated at room temperature for 5-20 minutes. Then in another tube, 2.4 mL of room temperature Opti-MEM, 80 µL of P3000, 13.3 µg of psPAX2, 4.5 µg of pMD2G, 1.8 µg of pAdVantage, and 19.5µg of the lenti vector of interest were combined. The contents of the two tubes were combined and incubated at room temperature for 30 minutes. This mixture was then added dropwise to a plate of Lenti-X cells and cultured overnight. The next morning the media was changed with 36 mL of fresh warm DMEM+Glutamax+10% FBS media supplemented with 72 µL viral boost reagent (ALSTEM, Cat. No. VB100). Then 2-3 days later the media containing virus was collected, centrifuged to remove cell debris and the supernatant was concentrated using Lenti-X concentrator, following the manufacturer's protocol (Takara Bio, Cat. No. 631231). The viral pellet was gently dissolved in PBS at either 1:10 or 1:100. Concentrated virus was then aliquoted and stored at -80°C for future use.

CRISPRi plasmids

The sgRNAs used in this study were cloned into either the pU6-sgRNA EF1Alpha-puro-T2A-BFP vector (gift from Jonathan Weissman; Addgene #60955)[24,44] or the pMK1334 CROPSeq vector (gift from Martin Kampmann; Addgene #127965)[24]. Guides were driven under the mouse U6 promoter. The sgRNA sequences were as follows: non-targeting controls: GTCCACCCTTATCTAGGCTA or GACCAGGATGGGCACCACCC; PARP1: GGGTGCGGCGTGTTCGGTGG; XRCC1: GGTAGAGTATGGGGTCCGAG. POLB sg1: GCGCCGGAGGGAGATCCCCA, POLB sg2: GCCAGCTTGA AGGAGGTACC, PNKP: GCCAGGGCTTGCCCGTCCGA.

Puromycin selection for CRISPRi knockdown i^{3}Neurons

To perform CRISPRi knockdown 1-4 million iPSCs were transduced with an aliquot of sgRNA-expressing virus immediately following an accutase split prior to cell attachment. Media was changed the next day to remove residual virus. Two days after transduction, iPSCs were split into accutase and plated at low density in E8+RI media containing 10 µg/mL puromycin. The following morning cells were washed with PBS and given fresh E8 or E8+RI media. Cells were then expanded for 1-2 days before inducing differentiation into neurons. sgRNA knockdown efficiency was tested at the iPSC stage and confirmed in day 7 i^{3}Neurons via QT-PCR.

Immunofluorescence and Microscopy

For imaging, i^{3}Neurons were plated onto 96-well plates (0.05x10^6 cells per well; Perkin Elmer, Cat. No. 6055302) or on µ-Slide glass bottom IBIDI slides (or 0.2x10^6 per well; IBIDI). Prior to fixation, if indicated, cells were treated with 0.1 mg/ml MMS for 15 min, 10 µM PARGi for 20 min, or 50 µM Etoposide (ETO) for 1h. Cells were then washed with PBS and fixed in 4% paraformaldehyde in PBS for 15 minutes at room temperature. Cells were then washed 3 times in PBS, permeabilized in 0.5% Triton X-100 for 5 minutes at room temperature and blocked in 1% BSA/ 0.2% Triton in PBS for 1 hour at room temperature before incubation with primary antibodies. Primary antibodies and dilutions used were as follows: anti-poly-ADP-ribose (1:500, Sigma-Aldrich MABE1031), anti-53BP1 (1:1000, Novus Biologicals, #NB100-305), anti-Phospho-Histone H2AX (1:5000, Millipore, #JBW301), and anti-tubulin β3 (TUBB3, 1:5000, Biolegend #801201). Immunofluorescence detection was achieved using fluorochrome-conjugated secondary antibodies as follows: Rhodamine Red-X AffiniPure Donkey anti-Mouse IgG (1:1000, Jackson Labs #715-295-151) for detecting of TUBB3; Alexa Fluor 488 goat anti-rabbit (1:1000, Invitrogen #A11034) for detection of PAR or 53BP1; Alexa Fluor 555 goat anti-mouse (1:2000, Invitrogen #A21422) for detection of γH2AX. EdU was visualized using Click-iT™ Plus EdU Cell Proliferation Kit for Imaging, Alexa Fluor™ 488 dye (Life Technologies, Cat. No. C10637) following the manufacturer instructions. Finally, DNA was counterstained with DAPI (2 µg/mL, Thermo Fisher Scientific #62248). Images were acquired on an inverted Nikon spinning-disk confocal microscope (Nikon Eclipse T1), using a 60x 1.40 NA oil-immersion objective.

Flow Cytometry

For cell cycle profiling, cells were incubated with 10 µM (5-ethynyl-2′-deoxyuridine) for 30 min at 37 °C and stained using the Click-IT EdU Alexa Fluor 488 or 647 Flow Cytometry Assay Kit (ThermoFisher) according to the manufacturer's instructions. DNA content was measured by DAPI (4′,6-diamidino-2-phenylindole, 0.5 µg ml^{-1}). Data analysis was done using FlowJo v10 software.

Rat primary neuron cell culture

All animal procedures were conducted following the NIH Guide for the Care and Use of Laboratory Animals, under the Animal Study Proposal #19-011 approved by the NICHD Animal Care and Use Committee. 10-11 week old pregnant albino rats were delivered to our facility on day 17 of gestation. They were housed under a 12 h light-dark cycle for 24 h with access to food and water ad libitum. The following day, the animals were sacrificed by carbon dioxide inhalation followed by decapitation prior to embryo extraction and preparation of neurons. Neurons from embryos of the same litter were pooled for each experiment.

Primary rat cortical neurons were prepared at embryonic day 18 (E18) as previously described[45]. Rats at E18 were harvested and euthanized by decapitation. The brain was collected and meninges were removed, after which cortexes were isolated in sterile Hank's medium (Hanks' Balanced Salt Solution (HBSS), 20 mM HEPES, pH 7.5). Cortexes were then collected and treated with 0.25% trypsin (Gibco), and 100 µg/ml DNase (Roche) for 15 minutes at 37°C. One volume of adhesion medium (Dulbecco's Modified Eagle Medium (DMEM) without phenol red, 4.5 g/L glucose, 25 mM HEPES, 10% heat-inactivated horse serum (Gibco), 100 U/mL penicillin and 100 mg/mL streptomycin) was added to stop trypsin enzymatic action. The tissue was then disrupted mechanically by pipetting it through a 10 ml serological pipet. Cells were then strained through a 70 µm nylon filter (Corning) and centrifuged at 700 g for 10 minutes. The cell pellet was resuspended in 5 ml of adhesion medium and cells counted. 10-25 million cells were plated on 10 cm culture dishes previously coated with poly-L-lysine (Sigma) and 5 µg/mL laminin (Roche). After 2 hours, the neurons were adherent to the plate and the medium was changed to complete neurobasal medium (CNB) (neurobasal medium (Gibco), 1X B27 serum-free (Gibco), 4.5 g/L glucose, and 100 U/mL, penicillin-streptomycin (Gibco)) and supplemented with 5µM aphidicolin (Aph) to eliminate residual dividing cells. Primary neurons neurons were then cultured for 7-8 days before harvesting for CHiP-seq or SAR-seq.

SAR-seq

Neurons and iMuscle cells were incubated with 20 µM EdU for 18 hours, unless otherwise noted. Cells were harvested and fixed as follows. Cells were washed with PBS, incubated with accutase for 5-10 mins, collected with a cell scrapper, pelleted at 200 × g for 5 minutes and resuspended in cold 0.1% BSA in PBS. Cold methanol was then added dropwise during slow vortexing to 80% final concentration. Samples were kept on ice for 20 minutes and then stored at -20°C until processing.

Copper catalyzes azide-alkyne click chemistry. For biotin labeling via Click-iT reaction, cells were first washed 1x in PBS, permeabilized with 0.2% Triton-X100/PBS for 10 minutes on ice, and then washed 1x in PBS. Then the following were added in order: 3 mM copper sulfate (Sigma), 50 µM biotin azide (ThermoFisher, Cat. No. B10184), and 1X Click-iT additive (ThermoFisher, Cat. No. C10424) for 2 hours shaking at room temperature. Cells were then washed one time in PBS and lysed in 50 mM Tris pH 8.0 with 1% SDS and Proteinase K overnight at 37°C. DNA was extracted with UltraPure Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v) (Invitrogen) according to manufacturer's instructions, followed by 2.5:1 volume ethanol and 1:10 volume sodium acetate precipitation. DNA pellets were resuspended in TE buffer and sheared to 150-200 bp fragments using Covaris S220 sonicator at 10% duty cycle, 175 peak incident power, 200 cycles per burst, for 240 seconds. DNA was again precipitated by 2.5:1 volume ethanol and 1:10 volume sodium acetate and resuspended in TE buffer. Biotin-EdU fragments were pulled down using MyOne Streptavidin C1 Beads (ThermoFisher, Cat. No. 650-01). Before pulldown, 35 µL of Dynabeads were washed two times with 1 mL 1X Wash and Binding buffer (1X W&B) (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 M NaCl, 0.1% Tween20) on a DynaMag-2 magnetic separator (Invitrogen, Cat. No. 12321D) and resuspended in 2X W&B (10 mM Tris-HCl pH8.0, 2 mM EDTA, 2 M NaCl, 0.2% Tween20). Equal volume of Dynabeads in 2X W&B were added to DNA in TE and incubated at 24°C shaking in a ThermoMixer C at 800 rpm for 30 minutes. Dynabeads bound to biotin-EdU fragments were washed three times in 1 mL of 1X W&B, two times in 1 mL EB, and one time in 1mL 1X T4 DNA Ligase Buffer (NEB). Dynabeads were resuspended in 50 µL end-repair reaction mix (1X T4 DNA Ligase Buffer, 0.4 mM of dNTPs, 2.7 U of T4 DNA polymerase (NEB), 9 U of T4 Polynucleotide Kinase (NEB), and 1 U of Klenow fragment (NEB)) and incubated at 24°C shaking at 800 rpm for 30 minutes. Dynabeads were washed one time in 1mL 1X W&B, two times in 1mL EB, and one time in 1 mL NEBuffer 2 (NEB) and resuspended in 50 µL of A-tailing reaction mix (1X NEB dA-Tailing Buffer and 20 U Klenow fragment exo- (NEB)), followed by incubation at 37°C shaking at 800 rpm for 30 minutes. Dynabeads were then washed again 1x in 1 mL NEBuffer 2 and resuspended in 115 µL of ligation reaction mix (1X Quick Ligase Buffer (NEB), 6000 U Quick Ligase (NEB), 5 nM annealed TruSeq truncated adapter) and incubated at 25°C shaking at 600 rpm for 20 minutes. Ligation reaction was stopped by adding 50 mM EDTA, and Dynabeads

were washed three times in 1 mL 1X W&B, three times in 1 mL EB, and finally resuspended in 8 µL EB + 10 µL 2X Kapa HiFI HotStart Ready Mix (Kapa Biosciences). 10 mM primers 5′-CAAGCAGAAGACGGCATACGA- GATXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T-3′ and 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCC CTACACGAC GCTCTTCCGATC*T-3′ (* indicates a phosphothiorate bond and a NNNNNN TruSeq index sequence) were added with 37 µL PCR reaction mix (20 µL 2X Kapa HiFI HotStart Ready Mix, 17 µL H_2O) for final volume of 60 µL. DNA was amplified using PCR program: 98°C, 45 seconds; 15 cycles [98°C, 15 seconds; 63°C, 30 seconds; 72°C, 30 seconds]; 72°C, 5 minutes. PCR products were separated from DynaBeads and cleaned using 1.8X volume AMPure Beads XP. 150-200 bp bands were isolated on 2% agarose gel and purified using QIA-quick Gel Extraction Kit (Qiagen). Prior to sequencing using Illumina NextSeq 550 (75 bp single read), library concentrations were calculated by KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems).

To specifically sequence only EdU-incorporated strands of DNA (Strand-Specific SAR-seq), SAR-seq was followed exactly as above with additional steps prior to PCR amplification. After washing three times in 1 mL 1X W&B and three times in 1 mL EB post-ligation, Dynabeads were washed in 50 µL 1X SSC buffer and resuspended in 20 µL of 0.15 M NaOH for 10 minutes at room temperature to denature DNA strands. Beads were placed back on the DynaMag-2 magnetic separator and washed one time with 20 µL 0.1 M NaOH, one time with 1 mL 1X W&B, two times with 1 mL EB, and resuspended in 8 µL EB + 10 µL 2X Kapa HiFI HotStart Ready Mix. Primers and PCR reaction mix were added as above.

In the indicated experiments, i^{3}Neurons were treated with the following compounds: Aphidicolin (Aph; 5 µM), etoposide (50 µM), olaparib (10 µM), velaparib (10 µM) or tazaparib (1 µM) were added along with EdU at 18 hours prior to harvest. POLAi (1 µM) or Aph (50 µM) was added with EdU 14 hours or 24 hours prior to harvest, respectively. Hydroxyurea (HU; 10 mM) powder was dissolved fresh into water to make a 1 M stock each time it was used. HU was added on day 3 of differentiation and again on day 6 at 18 hours prior to harvest along with EdU for a total of 4 days of treatment. For dideoxynucleoside (ddN) chain termination,

5 µM of each of ddA, ddT, ddG, and ddC chain-terminating nucleosides (20 µM total) were added to i^{3}Neuron culture for 18 hours prior to cell harvesting for END-seq and S1-END-seq experiments.

END-seq and S1-END-seq

To dissociate i^{3}Neurons for use in END-Seq, we performed a papain dissociation protocol modified from a previous protocol[46]. Papain (Worthington Biochemical Cat: LK003178) was dissolved into TrypLE™ Express Enzyme (1X) no phenol red and warmed at 37°C for 10 minutes. Then, day 7 i^{3}Neurons on a 15cm plate were washed with PBS and treated with 5mL papain/TRPLE for 1 minute at 37°C. Papain was removed with gentle pipetting and 5mL of trituration solution was added (30mL i^{3}Neuron Culture Media,10 µM ROCK inhibitor, and 1 vial of DNAse freshly dissolved). Cells were collected and gently pipetted 3-10 times in a conical tube using a wide-bore 10mL pipet, being careful not to over-digest the sample. Cells were then washed with PBS, pelleted and resuspended in PBS containing 0.1% BSA and 0.5mM EDTA, and kept on ice. Cells were processed for END-seq as described[47]. For S1-END-seq, cells were collected and embedded in 1% agarose plugs, lysed and digested with Proteinase K (1 hour at 50°C, followed by 7 hours at 37°C), washed with TE buffer, and then treated with RNAse A for 1 hour at 37°C. Plugs were then washed in EB and equilibrated in S1 nuclease buffer (40 mM sodium acetate pH 4.5, 300 mM NaCl, 2 mM $ZnSO_4$) for 30 minutes. 1.8 U of S1 nuclease was added to 100 µL of S1 nuclease buffer per plug and incubated on ice for 15 minutes to allow for the enzyme to diffuse into the plug. The reaction mix was then placed at 37°C for 20 minutes before addition of EDTA (10 mM final concentration) to terminate the reaction. Finally, plugs were processed through the standard END-seq protocol.

AsiSI induction in pre-B cells

Abelson-transformed murine pre-B cells[48] were retrovirally transduced with tetracycline-inducible ER-AsiSI[49]. Cells were arrested in G1 with 3 µM imatinib for 24 hours, followed by addition of 3 µg/mL doxycycline for 24 hours, and then further addition of 1 µM 4OHT for 18 hours to induce AsiSI nuclear localization, as previously described[19].

Construction of the Dox-Cas9-D10A nickase

Dox-inducible Cas9-D10A was constructed using isothermal assembly[50]. Briefly, a plasmid encoding Dox-inducible Cas9 nuclease was obtained from Addgene. pCW-Cas9-Blast was a gift from Mohan Babu (Addgene plasmid # 83481; http://n2t.net/addgene:83481; RRID: Addgene_83481). This plasmid was digested with NheI/BamHI and assembled with 2 PCR fragments (Nickase-P1, Nickase-P2) and transformed into competent cells. Gel extracted PCR fragments of Nickase-P1, Nickase-P2 were generated using the Q5 HotStart 2X mastermix with the primers listed below using Addgene-83481 as template.

Fragment	*Forward Primer*	*Reverse Primer*
Nickase-P1	GTCAGATCGCCTGGAGAATTG	tgC CAGGCCGATGCTGTACTTCT
Nickase-P2	AGAAGTACAGCATCGGCCTG Gca ATCGGCACCAACTCTGTGGG	TGCCTTGGAAAAGGCGCAAC

MCF10A Cas9D10A inducible cell line

To produce the MCF10A Cas9-D10A inducible cell line, we infected MCF10A cells (sourced from ATCC, not aunthenticated, and not mycoplasma tested) with lentivirus containing Dox-Cas9-D10A and cells were selected with 10 µg/mL blasticidin. Cas9-D10A expression was induced by 3 µg/mL doxycycline and confirmed by Western Blotting. Three guide RNAs (sequences: 5′-TGGGGCGTTTATCCGATGTC-3′; 5′-GCACTAGCCGGCCCGGACGT-3′; 5′- CCAGCCTGGTAGCGCCCCCA-3′) were cloned into Lenti-Guide-NLS–GFP vector[51] and the MCF10A Cas9-D10A inducible cell line containing the three guide RNAs were selected with 2 µg/mL puromycin. For identifying the nicks by S1-END-seq, cells were arrested in G1 for 48 hours with 5 µM Palbociclib with doxycycline added during the last 24 hours to induce Cas9D10A, followed by cell harvesting and S1-END-seq processing.

ChIP-seq and Western blotting

15 million i^{3}Neuron or rat neurons were fixed in 1% formaldehyde at 37°C for 10 minutes. The fixation reaction was quenched with glycine at a final concentration of 125 mM. Cells were spun down and washed twice with chilled PBS, and pellets were then snap frozen on dry ice and finally stored at -80°C until sonication. Sonication, immunoprecipitation, and library preparation steps were done as previously reported[17]. All antibodies were pre-conjugated to 40 µL of magnetic Protein A beads prior to immunoprecipitations: H3K4me1 (5 µg, Abcam #8895); MLL4 (antibody courtesy of Kai Ge); H3K27ac (5 µg, Abcam #4729); H3K27me3 (5 µg, Millipore #CS200603); H3K4me3 (6 µL Abcam #8580); H3K9me3 (10 µg, Active Motif, Cat. No. 39765); CTCF (6 µl, Millipore, Cat. No. 07-729); RNA Polymerase II (8 µg, Abcam #26721); H3K36me3 (5 µg, Abcam Cat. No. ab9050); anti-pan-ADP-ribose (5 µg, Millipore-Sigma MABE1016); XRCC1 (2.6 µg, Novus, Cat. No. NBP1-87154). For Western blotting, cells were collected and lysed in a buffer containing 50 mM Tris-HCl (pH 7.5), 200 mM NaCl, 5% Tween-20, 0.5% NP-40, 2 mM PMSF, 2.5 mM β-glycerophosphate (all from Sigma-Aldrich) and protease inhibitor cocktail tablet (complete Mini, Roche Diagnostics). Equal amounts of protein were loaded into precast mini-gels (Invitrogen) and resolved by SDS-PAGE. Proteins were blotted onto a nitrocellulose membrane, blocked with Intercept (TBS) blocking buffer (LI-COR Biosciences) and incubated with the corresponding primary/secondary antibodies: anti-DNA polymerase β (1:1000, Millipore #ABE1408), anti-Tubulin (1:10,000, Sigma-Aldrich #T5168), IRDye 800 CW goat

anti-rabbit (1:15,000, Li-Cor #926-32211), and IRDye 680 RD goat anti-mouse (1:15,000, Li-Cor #926-68070).

Selective chemical labeling and capture of 5hmC and 5fC

5hmC-Seal was performed as previously described[52] with modifications. Briefly, 80 µg genomic DNA was resuspended in TE buffer and sonicated to 200 bp fragments using a Covaris S220 sonicator. Fragmented DNA was precipitated by ethanol and sodium acetate and resuspended in TE buffer. The selective 5hmC chemical labeling was performed in 100 µl glucosylation buffer (50 mM HEPES buffer pH 8.0, 25 mM $MgCl_2$) containing above fragmented DNA, β-GT (NEB, Cat. No. M0357), UDP-Azide-Glucose (Active Motif, Cat. No. 55020), and incubated at 37°C for 1.5 hr. After the reaction, DNA was cleaned up with QIAquick Nucleotide Removal Kit. The labeled DNA was eluted with ddH_2O, after which 1mM DBCO-PEG4-Biotin (Click Chemistry Tools) was added and incubated at 37°C for 2 hours. Then, the biotin-labeled DNA was pulled down by C1 Streptavidin beads (ThermoFisher, Cat. No. 650-01) for 30 minutes at room temperature. The captured DNA fragments were processed for library construction as described in SAR-seq (above).

For 5fC-SEAL, we modified the previously described protocol[10], which reduces 5fC to 5hmC using $NaBH_4$[53]. We labeled the newly generated 5hmC (derived from 5fC reduction with $NaBH_4$) with an azide-modified glucose as described above for 5hmC-Seal. In brief, 80 µg of fragmented i^{3}Neurons genomic DNA was incubated in 100 µl glucosylation buffer (50 mM HEPES buffer pH 8.0, 25 mM $MgCl_2$) containing unmodified UDP-Glucose (NEB, Cat No. M0357), and β-GT for 1.5 hours at 37°C. The labeled DNA was cleaned up with QIAquick Nucleotide Removal Kit. Then, an equal volume of freshly prepared NaBH4 (Aldrich, Cat. No. 213462) solution was added to the glucose-blocked DNA solution. The reaction mixture was vortexed and incubated in a Thermomixer for 15 minutes at room temperature. The chemical labeling and capture were performed as described above for 5hmC-SEAL.

In situ Hi-C from hiPSC-derived i^{3}Neurons

Two *in situ* Hi-C libraries were generated from 10 million cultured hiPSC-derived i^{3}Neurons as described in[54]. Briefly, in situ Hi-C consists of 7 steps: (1) crosslinking cells with formaldehyde, (2) DNA digestion using MboI, (3) filling in and marking ends with Biotin, (4) proximity ligation, (5) DNA shearing, (6) pulling down the biotinylated ligation junctions with streptavidin beads, and (7) paired end sequencing. As quality control (QC), we confirmed efficient restriction, ligation and DNA shearing with an agarose DNA gel and for appropriate size selection in using Agilent 4200 TapeStation system after steps (5) and (6). For the final QC, we performed 100 paired end sequencing on the Illumina Nextseq to assess quality of the libraries based on percent of Intra-chromosomal reads, long Range (>20Kb) reads, and Library Complexity. The HiC libraries were sequenced on 150 paired end sequencing using Illumina Novaseq 6000.

RNA extraction, RNA-seq and quantitative real-time PCR (RT-PCR)

To extract RNA, cells were plated on 6-well dishes, washed with PBS, and then 500 µL of tri-reagent (Zymo research corporation, Cat. No. R2050-1-200) was added directly to the cells. The lysed cells were collected using a cell scraper. To isolate RNA, we used a Direct-zol RNA miniprep kit (Zymo Research Corporation, Cat. No. R2052), following manufacturer's instructions including the optional DNAse step. For QT-PCR, total RNA was reverse transcribed with iSCRIPT Advanced cDNA Synthesis Kit (Bio-Rad, Cat. No. 1725037) following manufacturer's instructions. The resulting cDNA was diluted 10-fold and used for quantitative real-time PCR (qRT-PCR) analyses with specific primer and probe sets (Bio-Rad, ACTB qHsaCEP0036280; PARP1 qHsaCEP0052423; XRCC1 qHsaCIP0033686; POLβ qHsaCEP0057881; FEN1 qHsaCEP0039485; PNKP qHsaCEP0057803 in a final volume of 20 µL, which contained 10 µL of SsoAdvanced™ Universal Probes Supermix (Biorad, Cat. No. 1725280), 3 µL of cDNA. qRT-PCR was performed in triplicate wells per sample on a CFX96 Real-Time System (Biorad). For RNA sequencing, six biological replicates were sequenced. Total RNA was enriched for polyA and sequenced 2x75 bp on a HiSeq machine.

ATAC-seq

ATAC-seq was performed as described previously[55]. i^{3}Neurons were grown on 96-well plates ($0.05x10^6$ cells per well). Cells were washed gently with PBS. Then, 100 µL of accutase per well was added and removed, and the plate was incubated at 37°C for 5 minutes. 50 µL cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL CA-630) was added directly to the well for 10 minutes. Cells were then pipetted 10-20 times to break clumps and centrifuged at 500 × g for 10 minutes at 4°C. The nuclei pellet was resuspended in the Nextera transposition reaction mix (25 µL 2x TD Buffer, 2.5 µL Nextera Tn5 transposase (Illumina, Cat. No. FC-121-1030), and 22.5 µL nuclease free H_2O) on ice, then incubated for 30 minutes at 37°C. The tagmented DNA was purified using the Qiagen MinElute kit and eluted with 10 µL Elution Buffer. Following purification, library fragments were amplified using the Nextera index kit (Illumina, Cat. No. FC-121-1011) under the following cycling conditions: 72°C for 5 minutes, 98°C for 30 seconds, followed by thermocycling at 98°C for 10 seconds, 63°C for 30 seconds, and 72°C for 1 minute for a total of five cycles. To prevent saturation due to over-amplification, a 5 µL aliquot was then removed and subjected to qPCR for 20 cycles for calculation of the optimal number of cycles needed for the 45 µL reaction that remained. The number of additional cycles required was determined by plotting linear Rn vs. Cycle and calculating the cycle number corresponding to a quarter of the maximum fluorescence intensity. Adding seven cycles to this estimate was found to yield optimal libraries. PCR reactions were subsequently cleaned with Agencourt AMPure XP beads (Beckman Coulter), run on a 2% agarose gel and a smear of 200-800 bp was cut and gel purified using QIAquick Gel Extraction Kit (QIAGEN). Library concentration was determined with KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems). Sequencing was performed on the Illumina Nextseq500 (75 bp paired-end reads).

Genome alignment

SAR-seq, END-seq, ChIP-seq and SEAL reads were aligned to the reference genome (hg19 for human i^{3}Neuron and iMuscle, mm10 for mouse pre B cells or rn6 for rat primary neurons) using bowtie (v1.1.2)[56] with parameters -n 3 -l 50 -k 1 for END-seq and -n 2 -l 50 -m 1 for the rest. ATAC-seq reads were aligned by bowtie (v2.4.1)[57]. RNA-seq reads were aligned by STAR (v2.7.6a)[58]. Functions "view" and "sort" of samtools (v1.11)[59] were used to convert and sort the aligned sam files to sorted bam files. Bam files were further converted to bed files by the bedtools (v2.29.2) bamToBed command[60]. Mitochondrial reads were removed in SAR-seq for intensity comparisons.

Peak calling

We used MACS (v1.4.3)[61] to call SAR-seq, XRCC1 ChIP-seq and ATAC-seq peaks. SAR-seq XRCC1, and CTCF ChIP-seq peaks with >10 fold-enrichment over background were kept. Peaks of hisone modification determined by ChIP-seq peaks were called by SICER[62] v2-1.0.2 with default parameters. Peaks within blacklisted regions (https://sites.google.com/site/anshulkundaje/projects/blacklists) were filtered[63]. Overlapped SAR-seq peaks from three non-treated replicates shown in Figure 1b were used for most of the analyses in this paper. As peaks of ddN S1-END-seq are always clustered, subpeaks were called by the PeakSplitter tool of PeakAnalyzer with parameters -c 15 -v 0.6[64]. Peak summits on both strands were identified, and the distance between peaks summit on the negative strand to its closed peak summit on the plus strand was calculated and represented in Fig 4d.

Quantification for sequencing data

For SAR-seq, SEAL, and ChIP-seq reads per kilobase of transcript, per million mapped reads (RPKM) was calculated. For RNA-seq, fragments

per kilobase of transcript per million mapped reads (FPKM) was calculated by cufflinks[41] (v2.2.1) based on the annotation from GENCODE v33[65]. Mean values of replicates were used for analyses.

Enhancer enrichment and super enhancer identification

Bedtools shuffle command with parameter (-chrom) was used to generate 1,000 random sets for SAR-seq peaks and ATAC-seq peaks, respectively, in order to estimate enrichments at genomic features. 1 kb regions upstream of transcription start sites from GENCODE v33 annotation were defined as promoters in analyses. Super enhancers were separated from conventional enhancers by identifying an inflection point of H3K27ac signal versus enhancer rank[66].

Gene Ontology (GO) analysis

We used the DAVID web-tool[67] to find the GO terms enriched for genes containing the top 2,000 SAR-seq peaks. The most significant Gene Ontology Biological Process terms and their associated FDR values were reported.

Motif analysis

MEME-ChIP[68] of MEME suite was used to identify the common sequence motif of the nucleotide sequences from ±500 bp around the summits for the top 5,000 SAR-seq peaks. The composite DNA sequence motif shown in Fig 4e was plot by ggseqlogo in R.

Kinetics Model for EdU labeling

The rate of EdU labeling was estimated under the assumptions that the fraction of a synthesis event at a given site is fixed with different EdU incorporation times and 18 hours is enough for EdU to incorporate all synthesis events at a given site. The fraction of EdU labeled synthesis event at *h* hours at a given site was denoted as $[EdU]_h=1-e^{-kh}$. *k* represents the rate of EdU labeling at a given site and $[EdU]_h$ was estimated by SAR-seq intensity at h hours divided by SAR-seq intensity at 18 hours. The intensity values (RPKM) of SAR-seq at different time points (1, 2, 4, 8 and 18 hours of EdU) was fitted to the function $[EdU]_h = 1-e^{-kh}$ using the nls package in R. The *k* value shown in Extended Data Fig. 2b was derived using the top 2,000 SAR-seq peaks.

Hi-C analysis

Using Juicer software[69], .hic files were generated, and normalized contact matrices and observed over expected normalized contact matrices were obtained with a dump command. Loop or domain calls were also done by juicer software[69], and interaction matrices were visualized by Juicebox software. After processing two replicates, final map was generated by merging these two replicates. The final bin resolution of the Hi-C map was 1 kb. From 2.1B raw reads generated by together with QC and data-generation runs, we obtained 1.5B final total contacts which are over Q30.

For compartment analysis, PC1 eigen vector values are extracted in 50k resolution by juicer software, then sign of eigen vector in each chromosome was adjusted according to the H3K27ac, H3K4me1 and ATAC-seq peaks distribution pattern in whole chromosome. Among H3K4me1 peaks which were not overlapped with TSS within 1kb, peaks overlapping with SARseq peaks were extracted for the analysis (SAR-seq+H3K4me1+). 10,000 random sequence were generated by bedtools random program with 2600 nt length which is the average length of H3K4me1 peaks. While 40% of random sequences are located in A compartment, 63% of SAR-seq +H3K4me1+ peaks were located in A compartment.

Visualization

BedGraph files were generated by bedtools genomecov, normalized by reads per million (RPM) and then converted to bigWig files using bedGraphToBigWig from UCSC pre-compiled utilities for visualization at UCSC genome browser[70,71].

Data matrices for heatmaps were calculated by computeMatrix and plotted by plotHeatmap of deepTools suite[72]. Venn diagrams were plotted by VennDiagram package in R. Confocal images were processed using FIJI. Schematics were created using BioRender, and figures compiled in Adobe Illustrator.

Quantification of Confocal Images

Confocal images were quantified using Nikon software. Images were first background subtracted. Then nuclei or EdU-positive cells were identified and counted using bright spot detection. Corresponding PAR signal intensity was measured per cell and reported as a mean. Data compilation and statistical analyses were performed using PRISM software.

Statistical Analysis

Statistical analysis was performed using R version 3.6.2 (http://www.r-project.org). The statistical tests are reported in the figure legend and main text.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The SARseq, S1-ENDseq, ChIPseq, data have been deposited in the Gene Expression Omnibus (GEO) database under the accession number GSE167259. iPSC enhancers were identified by H3K4me1 ChIP-seq from an ENCODE iPSC cell line (GSM2527632). Experimentally-validated enhancers was from VISTA genome browser for enhancers: https://enhancer.lbl.gov/.

40. Watanabe, S. *et al*. MyoD gene suppression by Oct4 is required for reprogramming in myoblasts to produce induced pluripotent stem cells. *Stem Cells* **29**, 505-516, https://doi.org/10.1002/stem.598 (2011).
41. Akiyama, T. *et al*. Efficient differentiation of human pluripotent stem cells into skeletal muscle cells by combining RNA-based MYOD1-expression and POU5F1-silencing. *Sci Rep* **8**, 1189, https://doi.org/10.1038/s41598-017-19114-y (2018).
42. Selvaraj, S. *et al*. Screening identifies small molecules that enhance the maturation of human pluripotent stem cell-derived myotubes. *Elife* **8**, https://doi.org/10.7554/eLife.47970 (2019).
43. Pawlowski, M. *et al*. Inducible and Deterministic Forward Programming of Human Pluripotent Stem Cells into Neurons, Skeletal Myocytes, and Oligodendrocytes. *Stem Cell Reports* **8**, 803-812, https://doi.org/10.1016/j.stemcr.2017.02.016 (2017).
44. Gilbert, L. A. *et al*. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647-661, https://doi.org/10.1016/j.cell.2014.09.029 (2014).
45. Farias, G. G., Britt, D. J. & Bonifacino, J. S. Imaging the Polarized Sorting of Proteins from the Golgi Complex in Live Neurons. *Methods Mol Biol* **1496**, 13-30, https://doi.org/10.1007/978-1-4939-6463-5_2 (2016).
46. Kirwan, P., Jura, M. & Merkle, F. T. Generation and Characterization of Functional Human Hypothalamic Neurons. *Curr Protoc Neurosci* **81**, 3 33 31-33 33 24, https://doi.org/10.1002/cpns.40 (2017).
47. Wong, N., John, S., Nussenzweig, A. & Canela, A. END-seq: An Unbiased, High-Resolution, and Genome-Wide Approach to Map DNA Double-Strand Breaks and Resection in Human Cells. *Methods Mol Biol* **2153**, 9-31, https://doi.org/10.1007/978-1-0716-0644-5_2 (2021).
48. Bredemeyer, A. L. *et al*. DNA double-strand breaks activate a multi-functional genetic program in developing lymphocytes. *Nature* **456**, 819-823, https://doi.org/10.1038/nature07392 (2008).
49. Santos, M. A. *et al*. DNA-damage-induced differentiation of leukaemic cells as an anti-cancer barrier. *Nature* **514**, 107-111, https://doi.org/10.1038/nature13483 (2014).
50. Gibson, D. G. *et al*. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345, https://doi.org/10.1038/nmeth.1318 (2009).
51. Noordermeer, S. M. *et al*. The shieldin complex mediates 53BP1-dependent DNA repair. *Nature* **560**, 117-121, https://doi.org/10.1038/s41586-018-0340-7 (2018).
52. Cui, X. L. *et al*. A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. *Nat Commun* **11**, 6161, https://doi.org/10.1038/s41467-020-20001-w (2020).
53. Dai, Q. & He, C. Syntheses of 5-formyl- and 5-carboxyl-dC containing DNA oligos as potential oxidation products of 5-hydroxymethylcytosine in DNA. *Org Lett* **13**, 3446-3449, https://doi.org/10.1021/ol201189n (2011).
54. Rao, S. S. *et al*. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680, https://doi.org/10.1016/j.cell.2014.11.021 (2014).
55. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218, https://doi.org/10.1038/nmeth.2688 (2013).

56. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25, https://doi.org/10.1186/gb-2009-10-3-r25 (2009).
57. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, https://doi.org/10.1038/nmeth.1923 (2012).
58. Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, https://doi.org/10.1093/bioinformatics/bts635 (2013).
59. Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, https://doi.org/10.1093/bioinformatics/btp352 (2009).
60. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, https://doi.org/10.1093/bioinformatics/btq033 (2010).
61. Zhang, Y. *et al*. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, https://doi.org/10.1186/gb-2008-9-9-r137 (2008).
62. Zang, C. *et al*. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952-1958, https://doi.org/10.1093/bioinformatics/btp340 (2009).
63. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**, 9354, https://doi.org/10.1038/s41598-019-45839-z (2019).
64. Salmon-Divon, M., Dvinge, H., Tammoja, K. & Bertone, P. PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. *BMC Bioinformatics* **11**, 415, https://doi.org/10.1186/1471-2105-11-415 (2010).
65. Harrow, J. *et al*. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760-1774, https://doi.org/10.1101/gr.135350.111 (2012).
66. Whyte, W. A. *et al*. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, https://doi.org/10.1016/j.cell.2013.03.035 (2013).
67. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, https://doi.org/10.1038/nprot.2008.211 (2009).
68. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696-1697, https://doi.org/10.1093/bioinformatics/btr189 (2011).
69. Durand, N. C. *et al*. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**, 99-101, https://doi.org/10.1016/j.cels.2015.07.012 (2016).
70. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204-2207, https://doi.org/10.1093/bioinformatics/btq351 (2010).
71. Kent, W. J. *et al*. The human genome browser at UCSC. *Genome Res* **12**, 996-1006, https://doi.org/10.1101/gr.229102 (2002).
72. Ramirez, F. *et al*. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165, https://doi.org/10.1093/nar/gkw257 (2016).

Acknowledgements We thank Anjana Rao, Samuel Wilson, Sergio Ruiz, Nuria Lopez-Bigas, David Wilson III, Yilun Sun, Yves Pommier, Marek Adamowicz, Paul Meltzer, and Kai Ge for helpful discussions and reagents; Raj Chari for constructing the CAS9D10A inducible nickase, and the CCR genomics core for help with sequencing. KWC is supported by Programme Grants from the UK Medical Research Council (MR/P010121/1), Cancer Research-UK (C6563/A7322), and by ERC Advanced Investigator (SIDSCA 694996) and a Royal Society Wolfson Research Merit Award. The M.W. laboratory is supported by the NINDS Intramural Research Program, the Chan Zuckerberg Initiative, and the Packard ALS Center. S.E.H. received funding from the BrightFocus Foundation. The A.N. laboratory is supported by the Intramural Research Program of the NIH, an Ellison Medical Foundation Senior Scholar in Aging Award (AG-SS-2633-11), the Department of Defense Awards (W81XWH-16-1-599 and W81XWH-19-1-0652) , the Alex's Lemonade Stand Foundation Award, and an NIH Intramural FLEX Award.

Author contributions W.W., S.E.H., W.J.N., J.P., K.W.C., M.E.W., and A.N. conceived, designed and analyzed the experiments. S.E.H., W.J.N., J.P., E.C., D.W. carried out experiments with assistance from K.S., J.C-M., N.V.W., D.Z., R.P., D.W., H-Y. S., S.C., M.P., R.P. and C.C., A.C. developed SAR-seq in the A.N. lab, W.J.N. developed S1-END-seq; W.W. designed bioinformatics pipelines and performed data analysis, and designed the figures. S. P. performed Hi-C, and S. K. J. and R.C. analyzed the Hi-C data; H.H., P.J.M and A.C provided insights. K.W.C., M.E.W., and A.N. wrote the paper with input from all co-authors. M.E.W. and A.N. supervised the study.

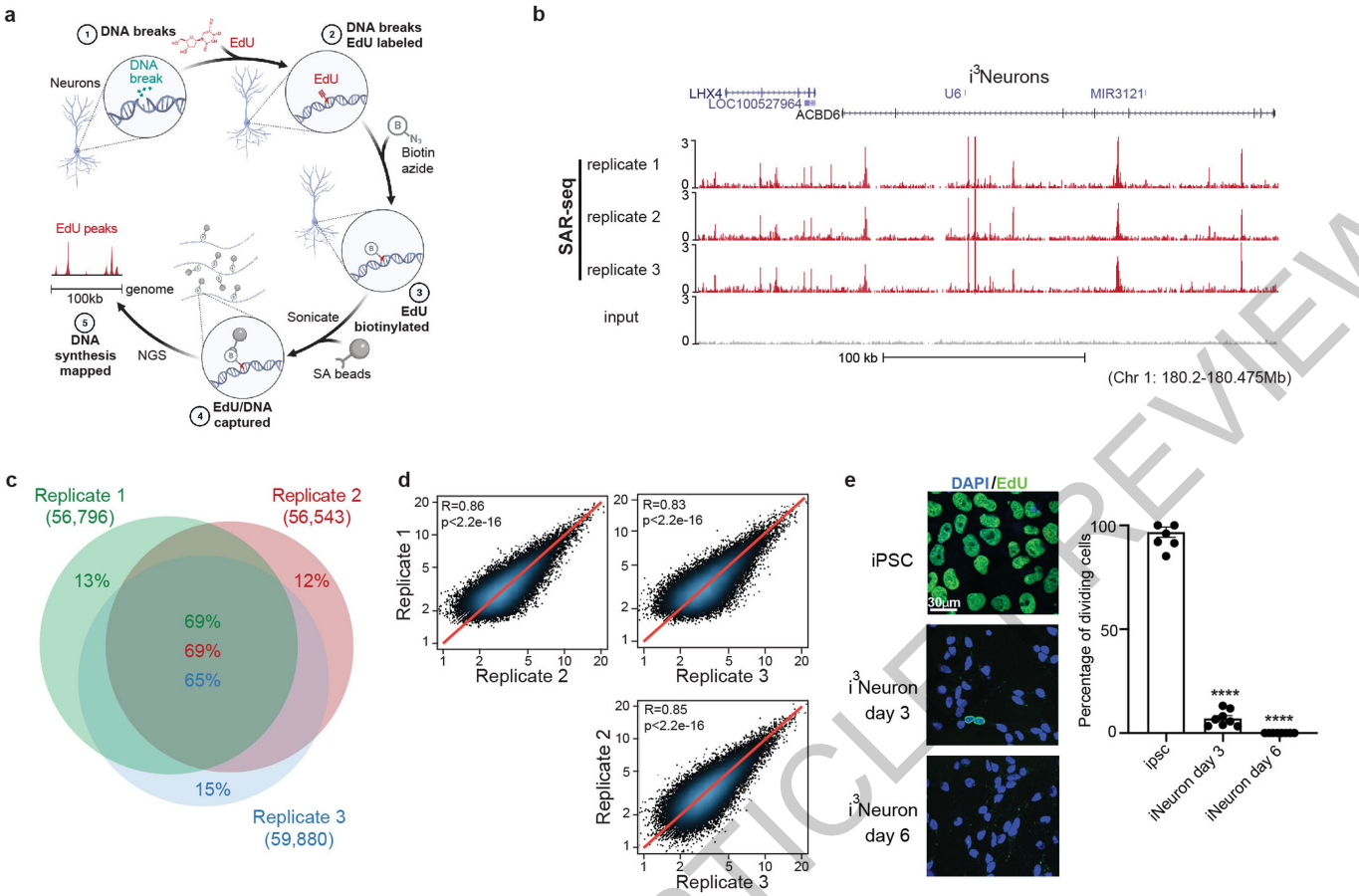
Competing interests The authors declare no competing interests.

Additional information

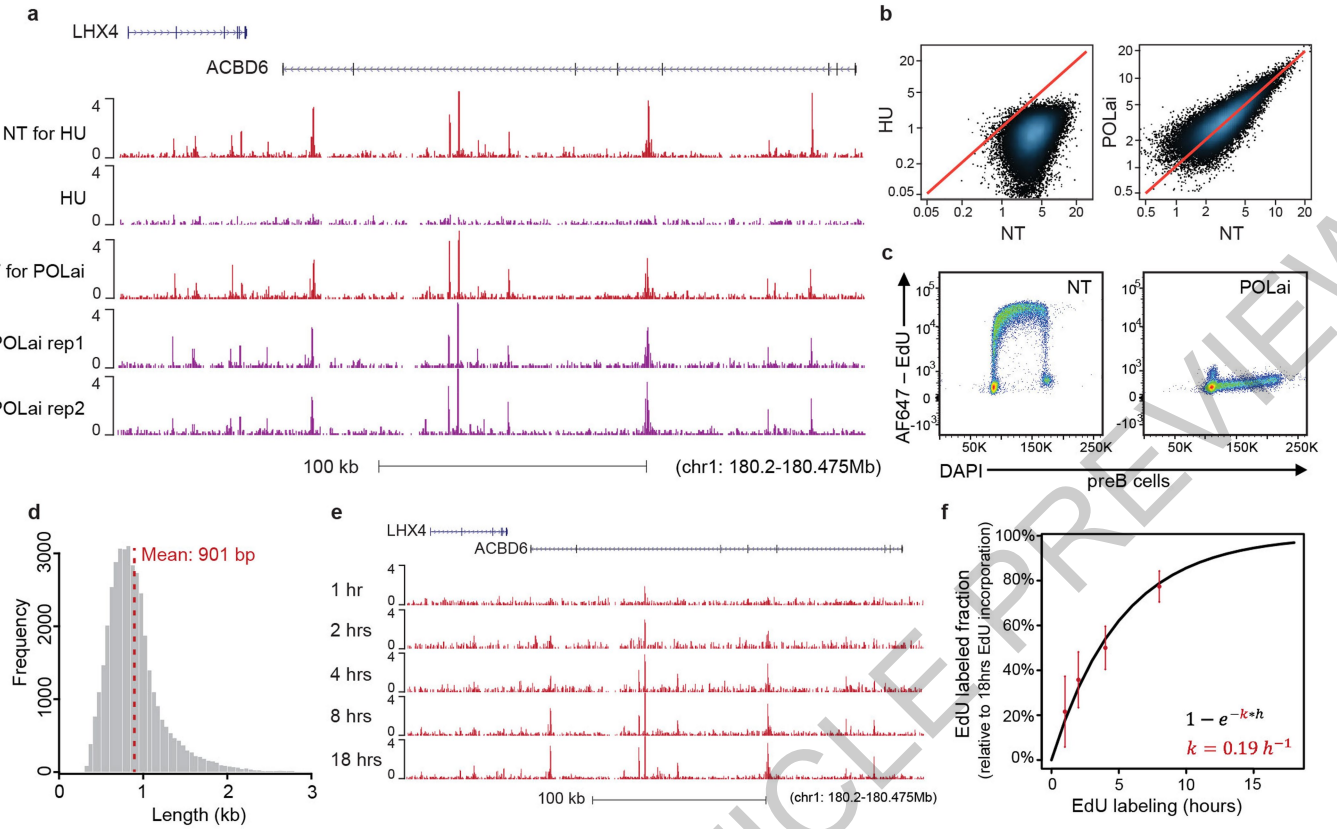
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03468-5.

Correspondence and requests for materials should be addressed to K.W.C., M.E.W. or A.N.

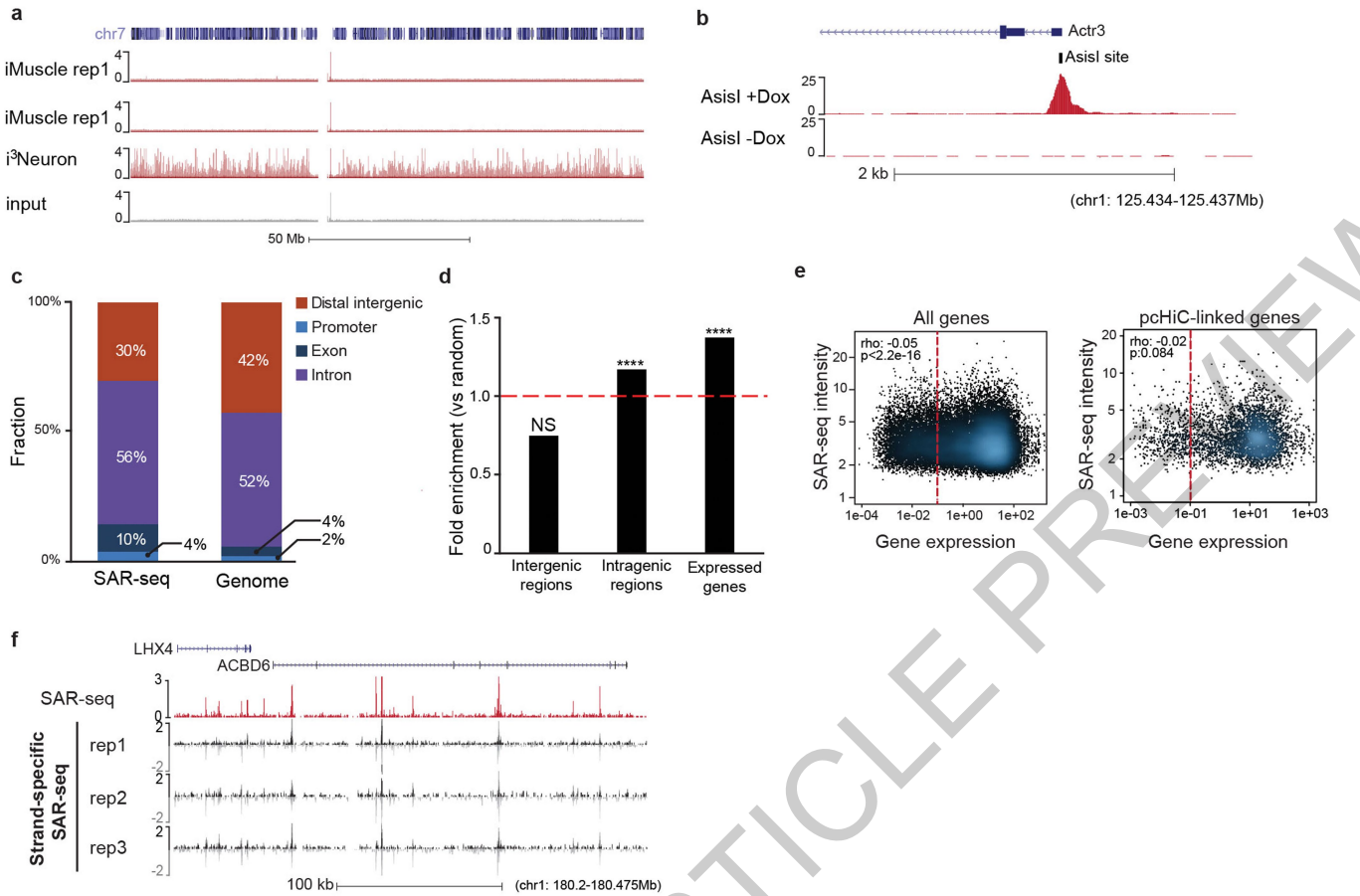
Peer review information *Nature* thanks Michael Weinfeld and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



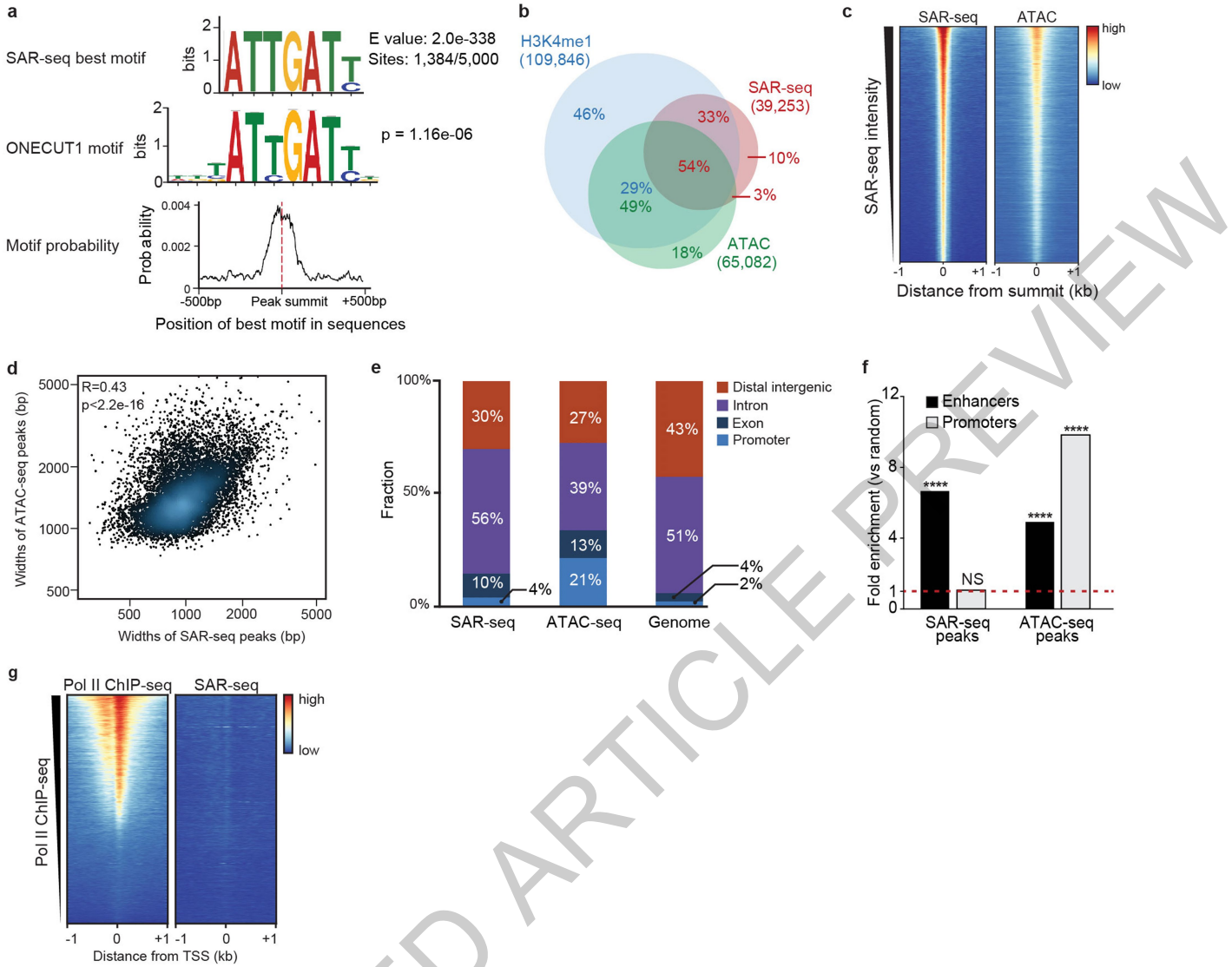
Extended Data Fig. 1 | Discrete genomic loci in neurons are associated with ongoing DNA synthesis. a) Schematic of SAR-seq (DNA synthesis associated with repair sequencing) methodology. Neurons grown in culture (*1*) are incubated with EdU to label sites of DNA repair synthesis (*2*). The incorporated genomic EdU is then conjugated to biotin via click chemistry (*3*), sheared by sonication to fragments of 150-200 bp and captured with streptavidin beads (*4*). Enriched DNA sequences are then PCR-amplified and subjected to next-generation sequencing (*5*). **b)** Genome browser screenshot displaying SAR-seq profiles as normalized read density (reads per million, RPM) for human iPSC-derived neurons (i^3Neurons). Three independent biological replicates are shown as well as input. Neurons were labeled with EdU for 18 hours and harvested on day 7 after induction of neuronal differentiation. All coordinates provided are from the hg19 reference genome for i^3Neurons. **c)** Venn diagram showing the overlap of SAR-seq peaks in i^3Neurons for three independent biological replicates. **d)** Scatter plots showing correlations of SAR-seq intensities (SAR-seq reads per kilobase per million mapped reads, RPKM) between three replicates in i^3Neurons. Pearson correlation coefficients and p values are indicated. **e)** Left panel: Representative images of EdU-biotin staining (green) showing cell proliferation in iPSCs, but not in post-mitotic i^3Neurons. i^3Neurons were treated with EdU on day 3 or day 6 and fixed on day 7. iPSCs were treated with EdU for 24h and fixed. Cells were counterstained with DAPI (blue). Note that different imaging conditions were used for iPSCs and i^3Neurons in the representative images. Right panel: quantification of EdU positive cells. Each dot represents the percent of dividing cells in one image (ipsc: n=8 images, n=397/410 cells EdU-positive; iNeuron day 3: n=8 images, n=35/483 cells EdU-positive; iNeuron day 7: n=8 images, n=0/523 cells EdU-positive). Data are presented as mean +/- SEM and are representative of three independent experiments.



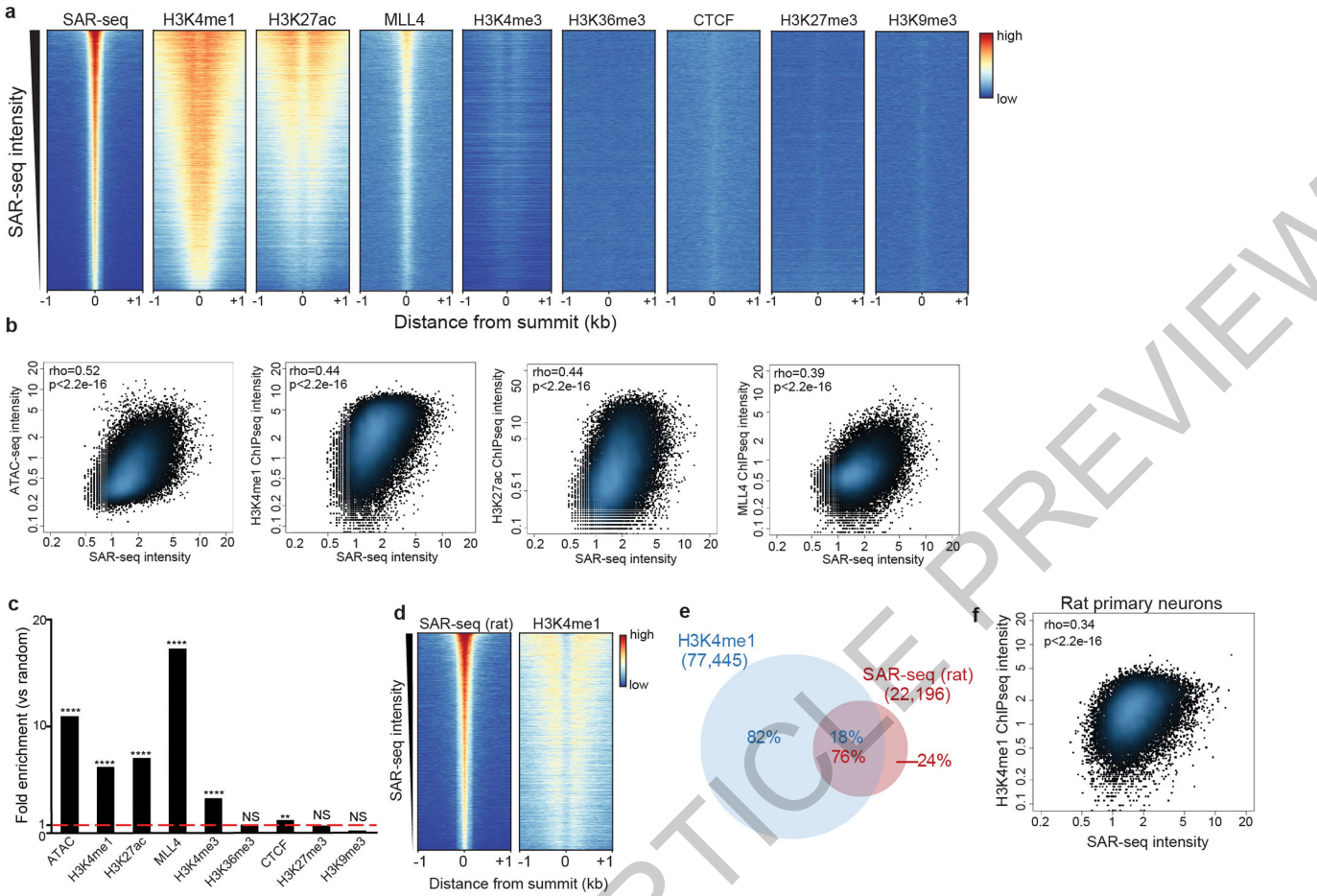
Extended Data Fig. 2 | Genomic characteristics of SAR-seq peaks. a) Genome browser screenshot showing SAR-seq in i^{3}Neurons treated with hydroxyurea (HU, n=1) or polymerase alpha inhibitor (POLαi, n=2). NT: non-treated. **b)** Scatter plots showing SAR-seq intensities (RPKM) for HU- (left) and POLαi- (right) treated compared to non-treated (NT) samples. **c)** Flow cytometry cell cycle profile of pre-B cells treated or non-treated (NT) with POLαi. Cells were pulsed with EdU for 30 minutes before collecting cells for flow cytometry. Cells were counterstained with DAPI (n=1). For an example of gating strategy used for flow cytometry in **c** and Extended Data 9b, please see Supplementary Figure 1. **d)** Histogram of individual SAR-seq peak widths, revealing an average peak width of 901 bp. **e)** Genome browser screenshot showing SAR-seq in i^{3}Neurons harvested after 1 hour, 2 hours, 4 hours, 8 hours, or 18 hours of EdU incubation (n=2 for each). **f)** Graph showing the fraction of EdU labeling in i^{3}Neurons (relative to maximum labeling at 18 hours) as a function of time calculated from top 2000 peaks. Red points and error bars represent mean and standard deviation of the relative levels of EdU measured from experimental data. Black line represents the theoretical model after fitting, with k being the rate of EdU labeling.



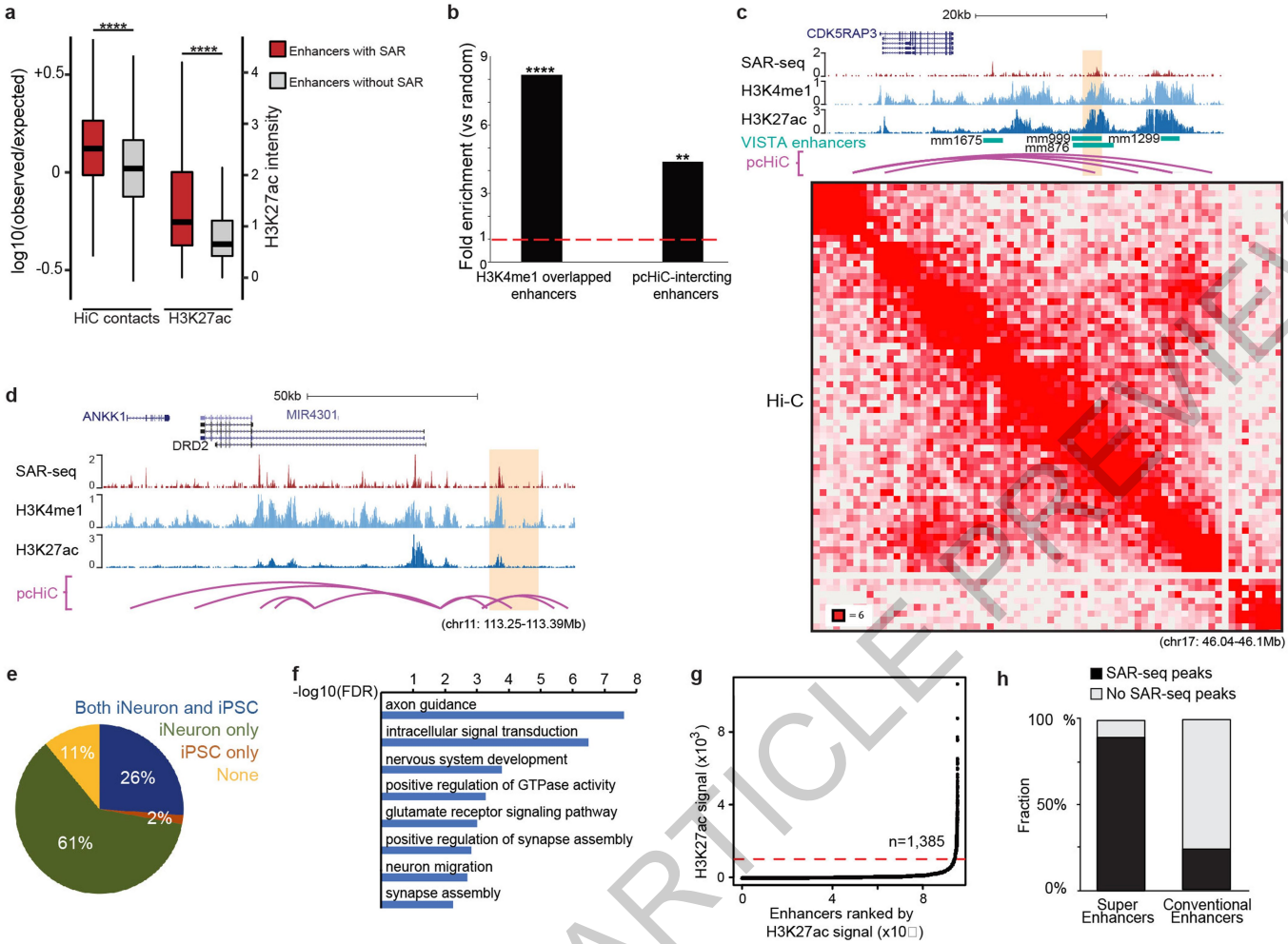
Extended Data Fig. 3 | SAR-seq enrichment at neuronal intragenic regions. **a)** Genome browser screenshot of chromosome 7 showing lack of localized DNA synthesis in two independent biological replicates of SAR-seq performed in iMuscle cells (n=2) incubated with EdU for 18 hours compared to SAR-seq in i^{3}Neurons as well as input DNA. **b)** Genome browser screenshot displaying SAR-seq peak at a representative AsiS1 restriction enzyme site (tick mark). AsiS1 expression was induced for 18 hours (+Dox, n=1) vs non-treated (-Dox, n=1) in G0-arrested, Abelson virus-transformed murine pre-B cells as described[19]. **c)** Distribution of SAR-seq peaks with respect to different genomic features compared to genome-wide distribution of the hg19 human reference genome. Promoters are defined as 1kb upstream of transcription start sites. Distal intergenic represents promoter-excluded intergenic regions. **d)** Fold enrichment of SAR-seq peaks in intergenic regions, intragenic regions and expressed genes compared to 1000 sets of randomly shuffled regions of the same sizes and chromosome distribution (one-sided Fisher's Exact test, ****p <2.2e^{-16}, NS: not significant). **e)** Left panel: Scatterplot showing correlation of SAR-seq intensity (RPKM) with transcript level of genes containing SAR-seq peaks measured by RNA-seq (n=3) (Fragments Per Kilobase of transcript per Million mapped reads, FPKM) in i^{3}Neurons. 71% of SAR-seq peaks are at expressed genes (FPKM≥0.1; red dashed line: FPKM=0.1). Right panel: correlation of SAR-seq intensity with transcript levels of linked genes determined by pcHiC in i^{3}Neurons (red dashed line: FPKM=0.1). Spearman correlation coefficients and p values are indicated. **f)** Genome browser screenshot comparing SAR-seq vs. strand-specific SAR-seq (n=3) that discriminates which strand is labeled with EdU in i^{3}Neurons. Both strands show labeling in three biological replicates. Strand-specific SAR-seq reads are separated by positive (black) and negative (grey) strands.



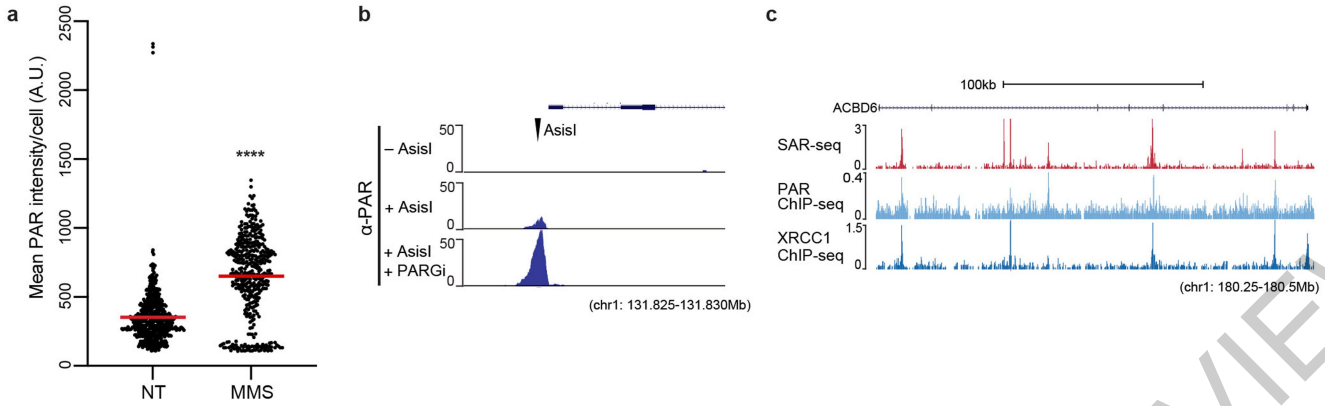
Extended Data Fig. 4 | Motif discovery of SAR peaks and comparison with ATAC-seq peaks. a) Motif analysis for sequences within ± 500 bp surrounding the summit of the top 5,000 SAR-seq peaks in i^{3}Neurons. Upper panel: the best motif discovered by the MEME suite. 1,384 out of 5,000 sites have this motif. Middle panel: TOMTOM motif tool used to compare SAR-seq motif shown above with databases of known motifs. The transcription factor ONECUT1 was identified as the most similar motif and its consensus sequence is shown. P value for motif comparison with ONECUT1 motif is indicated. Bottom panel: position distribution of the best motif (upper panel) within ± 500 bp of the SAR-seq peak summit. The best motif is centered on the SAR-seq peak summit. **b)** Venn diagram illustrating the overlap between H3K4me1 ChIP-seq, ATAC-seq and SAR-seq peaks in i^{3}Neurons. Statistical significance of the overlaps between SAR-seq, H3K4me1 ChIP-seq and ATAC-seq peaks was determined using randomly shuffled datasets (N=1,000) by one-sided Fisher's Exact test (the p value for overlap between H3K4me1 ChIP-seq/SAR-seq peaks is $p<2.2e^{-16}$, and for ATAC-seq/SAR-seq peaks is $p<2.2e^{-16}$). Fraction of different overlapping groups are labeled in red for SAR-seq peaks, green for ATAC-seq peaks and blue for H3K4me1 ChIP-seq peaks. **c)** Heatmaps of SAR-seq and ATAC-seq signal ±1kb around SAR-seq peak summits in i^{3}Neurons, ordered by SAR-seq intensity. **d)** Scatter plot comparing widths of ATAC-seq peaks and SAR-seq peaks for top 10000 overlapped peaks in i^{3}Neurons. Pearson correlation coefficient and p value are indicated. **e)** Distribution of SAR-seq and ATAC-seq peaks with respect to different genomic features compared to genome-wide distribution on the hg19 human reference genome. Promoters are defined as 1 kb upstream of transcription start sites and distal intergenic represents promoter-excluded intergenic regions. **f)** Bar graph showing the fold enrichment of SAR-seq and ATAC-seq peaks located at enhancers (black) and promoters (grey) compared to 1000 sets of randomly shuffled regions of the same sizes and chromosome distributions, respectively (one-sided Fisher's Exact test, ****$p<2.2e^{-16}$, NS: p=0.0783, not significant). **g)** Heatmap of RNA Polymerase II (Pol II, n=1) ChIP-seq and SAR-seq in i^{3}Neurons ±1kb surrounding the transcription start site (TSS) in i^{3}Neurons, ordered by Pol II ChIP-seq intensity.



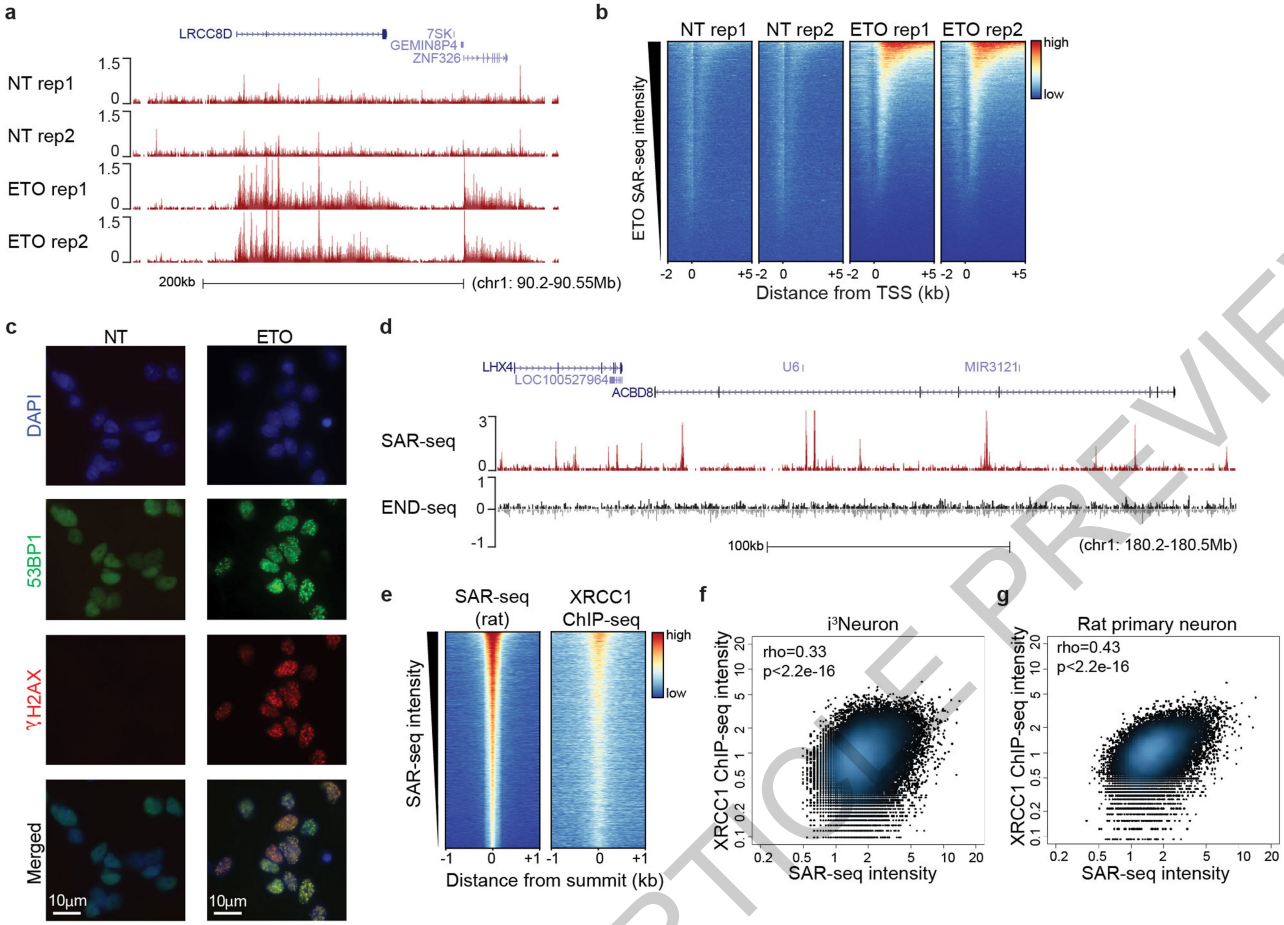
Extended Data Fig. 5 | The correlation between SAR-seq and chromatin features. a) Heatmaps of SAR-seq and ChIP-seqs of enhancer markers (H3K4me1, H3K27ac and MLL4), other chromatin markers at accessible regions (H3K4me3, H3K36me3 and CTCF) and chromatin silencing markers (H3K27me3 and H3K9me3) ±1kb around the SAR-seq peak summit in i^3Neurons, ordered by SAR-seq intensity. **b)** Scatter plots showing the correlation between SAR-seq intensity and ATAC-seq, H3K4me1, H3K27ac and MLL4 ChIP-seq intensities (RPKM) ±1 kb around the SAR-seq peak summits in i^3Neurons. Spearman correlation coefficients and p values are indicated. **c)** Fold enrichment of SAR-seq peaks at ATAC-seq peaks, ChIP-seq peaks of enhancer related marks (H3K4me1, H3K27ac and MLL4), additional chromatin marks at accessible regions (H3K4me3, H3K36me3 and CTCF) (all n=1) and chromatin silencing marks (H3K27me3 and H3K9me3) (both n=1) in i^3Neurons. N=1,000 randomly shuffled datasets were generated to test the significance using one-sided Fisher's Exact test ($p<2.2e^{-16}$ for H3K4me1, H3K27ac, ATAC-seq and MLL4, $p=1.85e^{-316}$ for H3K4me3, p=0.00116 for CTCF, **p<0.001, ****p <0.00001, NS: not significant). **d**) Heatmaps of SAR-seq and H3K4me1 ChIP-seq signal ±1kb around the SAR-seq peak summit in primary rat cortical neurons, ordered by SAR-seq intensity. **e)** Venn diagram showing the overlap between H3K4me1 and SAR-seq peaks in rat primary neurons. N=1,000 randomly shuffled datasets were generated to test the significance using one-sided Fisher's Exact test: $p<2.2e^{-16}$. **f)** Scatter plot showing the correlation between SAR-seq and H3K4me1 ChIP-seq intensities (RPKM) ±1kb around the SAR-seq peak summits in rat primary neurons. Spearman correlation coefficient and p values are indicated.



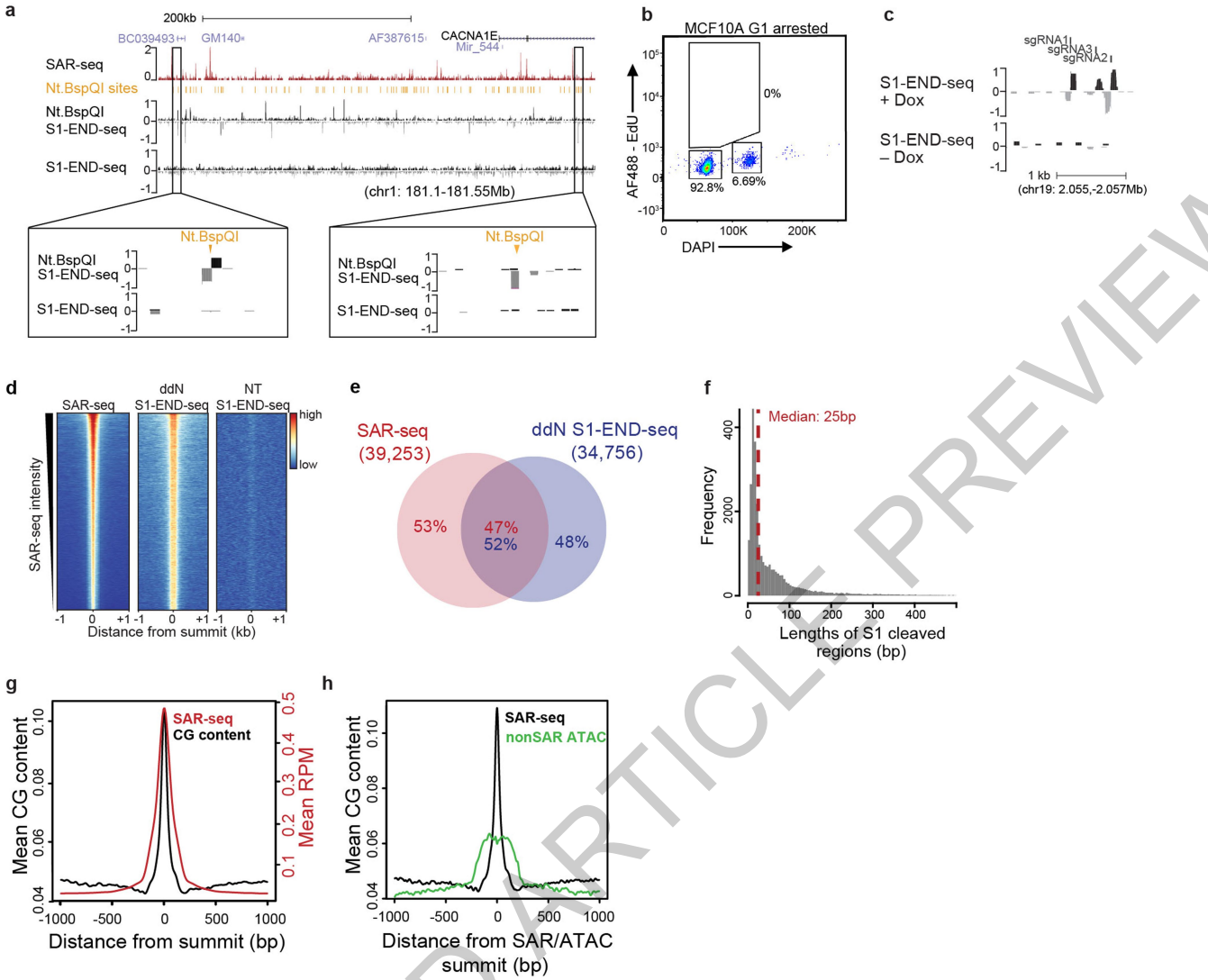
Extended Data Fig. 6 | SAR-seq enrichment at neuronal enhancers. **a)** Boxplot showing Hi-C contacts (left, n=4) and H3K27ac (right, n=1) levels at enhancers with SAR-seq peaks (red) or without SAR-seq peaks (grey). Contacts were defined as Hi-C interactions between an H3K27ac+ enhancer (with and without SAR) and its closest promoter within the TAD domain. For comparative purposes observed contacts were normalized to expected contacts. The top, centre mark, and bottom hinges of the box plots, respectively, indicate the 75th, median, and 25th percentile values. Statistical significance was determined using one sided Wilcoxon rank sum test. P is $1.64e^{-225}$ for HiC contacts and $p<2.2e^{-16}$ for H3K27ac intensity (****$p<0.00001$). **b)** Fold enrichment of SAR-seq peaks at *in vivo* validated enhancers from VISTA Enhancer Browser database that overlap with H3K4me1 ChIP-seq peaks (left, $p=1.42e^{-53}$) or at promoter interacting regions determined by pcHiC (right, $p=1.574e^{-09}$). N=1,000 randomly shuffled datasets were generated to test the significance using one-sided Fisher's Exact test (**$p<0.001$, ****$p<0.00001$). **c)** Genome browser screenshots showing SAR-seq, H3K4me1 and H3K27ac ChIP-seq, and pcHiC and Hi-C profiles at representative enhancers (highlighted in orange) interacting with CDK5RAP3 promoter and **(d)** the DRD2 promoter. Both enhancers have been validated to promote transcription of their respective genes using CRIPSR techniques in i^{3}Neurons[10]. The CDK5RAP3 enhancer also overlaps with *in vivo* validated enhancers from VISTA Enhancer Browser database. In the Hi-C contact matrix the intensity of each pixel represents the normalized number of contacts between a pair of loci. The maximum intensity is indicated at the lower left corner. **e)** Pie-chart showing distribution of i^{3}Neuron SAR-seq peaks in iPSC-specific, i^{3}Neuron- specific and shared iPSC- and i^{3}Neuron-enhancers. Approximately 56 million and 49 million single end reads were sequenced for the H3K4me1 ChIP-seq in iPSC and i^{3}Neurons, respectively, with approximately 100,000 peaks called in both cell types. **f)** Top biological processes enriched for the genes containing the 2,000 most intense SAR-seq peaks determined by Gene Ontology (GO) analysis. The x-axis represents the enrichment value as the logarithm of False Discovery Rates (FDR). **g)** H3K27ac signal at enhancers in i^{3}Neurons ranked by H3K27ac ChIP-seq intensity. Red dashed line indicates the inflection point of H3K27ac signal used to determine super enhancers (cutoff: 1000). Accordingly, 1,385 enhancers were defined as super enhancers. **h)** Bar graph showing the fraction of super-enhancers (left) and conventional enhancers (right) that overlap with SAR-seq peaks. The super-enhancers in the i^{3}Neurons were defined by H3K27ac ChIP-seq intensity in **g**.



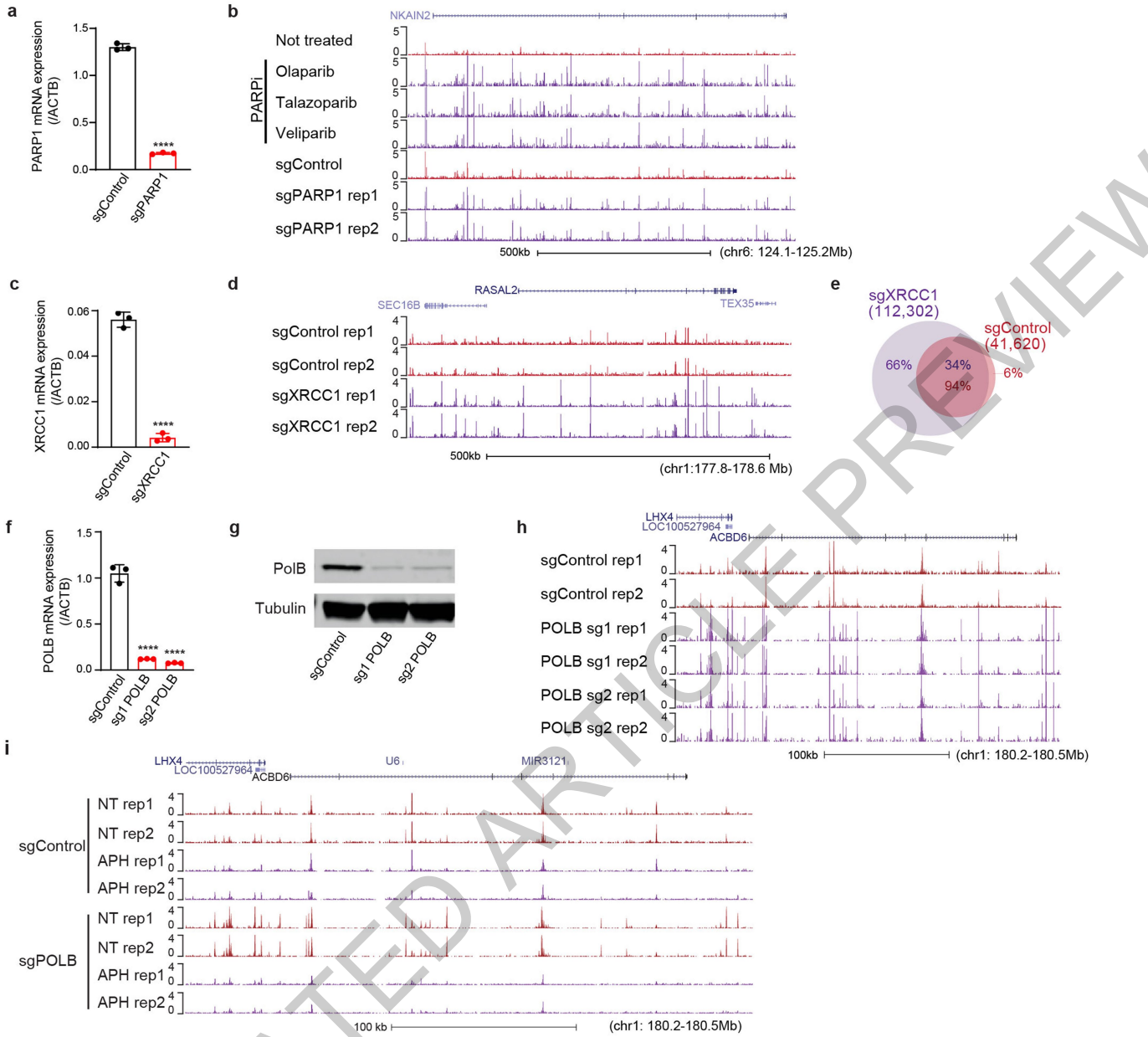
Extended Data Fig. 7 | PARP and XRCC1 recruited to DNA repair sites. **a)** Quantification of PAR levels with (n=422 cells) and without (n=541 cells) MMS treatment (red line: mean value). Each dot represents one cell. Statistical significance was determined using two sided Mann-Whitney test (**** $p<0.0001$). Data are representative of three independent experiments. **b)** Anti-ADP-ribose ChIP-seq signal (n=1) at an AsiSI restriction enzyme cut site (tick mark) in Abelson virus-transformed murine pre-B cells. Cells were arrested in G0, and AsiSI double-strand breaks were induced for 18 hours prior to ChIP. Note that ADP-ribose is enriched at cleaved AsiSI sites and is increased by 20 min treatment with PARGi prior to fixation (AsiSI + PARGi), which is indicative of the presence of poly(ADP-ribose). **c)** Genome browser screenshot illustrating the overlap between SAR-seq (n=3), PAR (n=1) and XRCC1 (n=1) ChIP-seq signals in i^3Neurons. Cells for PAR ChIP-seq were incubated with PARGi for 20 minutes prior to fixation.



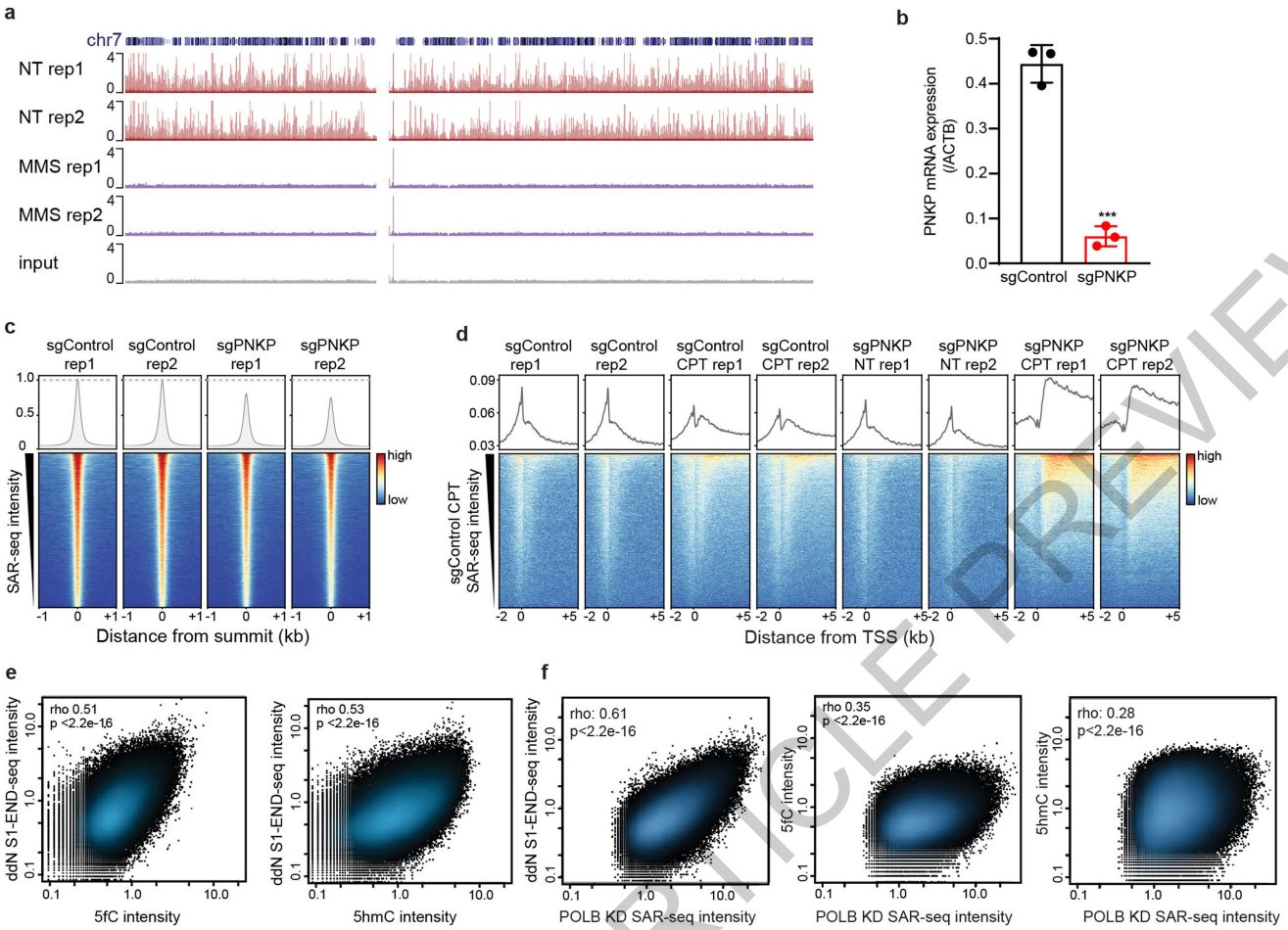
Extended Data Fig. 8 | Mapping regions of DNA damage and repair in neurons. a) Genome browser example of SAR-seq profiles in non-treated (NT) or etoposide (ETO)-treated (18 hours, 50μM) i^{3}Neurons. Data are from two biological replicates. **b)** Heatmaps for SAR-seq in nontreated (NT) or etoposide- (ETO) treated (18 hours 50μM) i^{3}Neurons at -2kb to +5kb of the transcription start sites (TSS) ordered by ETO SAR-seq intensity. **c)** Immunofluorescence staining of DSB markers γ-H2AX (red) and 53BP1 (green) in non-treated or one-hour ETO treated i^{3}Neurons. Data are representative of three independent experiments. **d)** Genome browser showing SAR-seq and END-seq profiles in non-treated i^{3}Neurons. Note that END-seq, which detects DSBs specifically[19], does not detect any enriched signal (i.e. above background) at SAR-seq peaks. END-seq signals are separated by positive (black) and negative (grey) strands. END-seq data are representative of two independent experiments. **e)** Heatmaps of SAR-seq and XRCC1 ChIP-seq (n=1) ±1kb around SAR-seq peak summits in cultured rat primary neurons, ordered by SAR-seq intensity. **f)** Scatter plots showing the correlation between SAR-seq and XRCC1 ChIP-seq intensities (RPKM) ±1kb around the SAR-seq peak summits in i3Neuron or rat primary neurons **(g)**, respectively. Spearman correlation coefficients and p values are indicated.



Extended Data Fig. 9 | S1-END-seq mapping of SSBs. a) Genome browser screenshot showing profiles of SAR-seq and S1-END-seq. Agarose plugs were incubated with or without the restriction enzyme Nt.BspQI prior to S1 treatment (n=1). Zoomed-in view of Nt.BspQI sites (tick mark) display S1-END-seq detection upon Nt.BspQI treatment. S1-END-seq reads are separated by positive (black) and negative (grey) strands. **b)** Flow cytometry profile of G1-arrested MCF10A cells pulsed with EdU. For G1 arrest, MCF10A cells were treated with Palbociclib (1μM) for 48 hours to arrest cells in G1. Doxycycline was added in the last 24 hours to induce nickase expression. Data representative of three independent experiments. For example of gating strategy used for flow cytometry in **b** and Extended Data 2c, please see Supplementary Figure 1. **c)** Genome browser screenshot showing S1-END-seq profiles at three Cas9 nickase targeting sites (tick marks: sgRNAs 1-3) in the G1-arrested MCF10A cells treated with doxycycline (+Dox) to induce Cas9 expression (n=1). S1-END-seq signals are separated by positive (black) and negative (grey) strands. **d)** Heatmaps of SAR-seq and S1-END-seq signal in i^{3}Neurons with ddN or without incubation (NT) ±1 kb surrounding the SAR-seq peak summits, ordered by SAR-seq intensity. **e)** Venn diagram showing the overlap between S1-END-seq peaks incubated with ddN and SAR-seq peaks in i^{3}Neurons. N=1,000 randomly shuffled datasets were generated to test the significance using one-sided Fisher's Exact test: p <2.2e-16. **f)** Distribution of the size of the gaps between positive- and negative-strand S1-END-seq peak summits in i^{3}Neurons incubated with ddN. The median gap size is 25 bp (red dashed line). Positive-strand peak represents the right end and negative-strand peak represents the left end of a detected DSB. **g)** Aggregate plots showing the distribution of CG dinucleotides (black) at ±1kb surrounding SAR-seq peak summits overlaid with SAR-seq signal (red). **h)** Aggregate plots showing the distribution of CG dinucleotides (black) at ±1kb around SAR-seq peak summits or summits of ATAC-seq peaks (green) that are H3K4me1 positive but do not overlap with SAR-seq.



Extended Data Fig. 10 | PARP, XRCC1 or POLB deficiency increases SAR. **a)** Quantitative RT-PCR analysis showing PARP1 mRNA transcript level in i^{3}Neurons after CRISPRi knockdown (sgControl: control non-targeting sgRNA; sgPARP1: an sgRNA targeting PARP1), cultured in parallel with samples used for SAR-seq. p is 7.72e^{-07} by unpaired two-tailed Student's t test, ****: p <0.00001. (n=3). **b)** Genome browser screenshot displaying SAR-seq profiles from two biological replicates of i^{3}Neurons treated with PARP inhibitors olaparib, talazoparib, veliparib, or CRISPRi-mediated knockdown with a control non-targeting sgRNA (sgControl) or an sgRNA targeting PARP1 (sgPARP1), in duplicates. NT: non-treated. **c)** Quantitative RT-PCR analysis showing XRCC1 mRNA transcript level in i^{3}Neurons after CRISPRi knockdown (sgControl: control non-targeting sgRNA; sgXRCC1: an sgRNA targeting XRCC1), cultured in parallel with samples used for SAR-seq. p is 1.88e^{-05} by unpaired two-tailed Student's t test, ****: p <0.00001. (n=3). **d)** Genome browser screenshots of SAR-seq profiles in i^{3}Neurons expressing CRISPRi non-targeting sgRNAs (sgControl) or targeting XRCC1 (sgXRCC1), in duplicate. **e)** Venn diagram showing the overlap of SAR-seq peaks between i^{3}Neurons expressing non-targeting sgRNA (sgControl) or targeting XRCC1 (sgXRCC1). N=1,000 random datasets were generated to test significance of overlap (one-sided Fisher's Exact test: p<2.2e^{-16}). **f)** Quantitative RT-PCR analysis showing POLβ mRNA transcript levels in i^{3}Neurons after CRISPRi knockdown (sgControl: control non-targeting sgRNA; POLB sg1 or sg2: two independent sgRNAs targeting POLβ), cultured in parallel with samples used for SAR-seq. p is 6.98e^{-05} for sg1 and 5.82e^{-05} for sg2 by unpaired two-tailed Student's t test, ****: p <0.00001. (n=3). **g)** Western blot showing POLB protein levels in i^{3}Neurons after CRISPRi knockdown (sgControl: control non-targeting sgRNA; POLB sg1 or sg2: two independent sgRNAs targeting POLB), cultured in parallel with samples used for SAR-seq (n=1). For gel source data, see Supplementary Fig. 2. **h)** Genome browser screenshots of SAR-seq profiles from two biological replicates of i^{3}Neurons expressing CRISPRi non-targeting sgRNAs (sgControl) or targeting POLβ (sgPOLB), in duplicates. **i)** Genome browser screenshots of SAR-seq profiles from two biological replicates of i^{3}Neurons expressing CRISPRi non-targeting sgRNAs (sgControl) or targeting POLβ (sgPOLβ). Cells were pre-treated or non-treated (NT) with 50 μM Aph for 24 hours, and then also during incubation with EdU.



Extended Data Fig. 11 | Localized SSB repair in neurons correlates with sites of oxidized 5-methylcytosine. a) Genome browser screenshot of chromosome 7 showing SAR-seq profiles from two biological replicates and input DNA in i^{3}Neurons without MMS treatment (NT, n=2) or after treatment with 0.1mg/ml MMS (n=2) for the final 15 min of an 18 hr incubation with EdU. After Streptavidin pull-down and PCR amplification, total DNA was quantified: NT rep1: 0.95 µg; NT rep2: 1.7 µg; MMS rep1: 3.8 µg; MMS rep2: 4.5 µg. Stochastic DNA damage results in loss of DNA synthesis at recurrent sites. **b)** Quantitative RT-PCR analysis showing PNKP mRNA transcript level in i^{3}Neurons after CRISPRi knockdown (sgControl: control non-targeting sgRNA; sg PNKP: an sgRNAs targeting PNKP), cultured in parallel with samples used for SAR-seq. p is 0.00015 by unpaired two-tailed Student's t test, ***: p <0.0001. (n=3). **c)** Heatmaps of SAR-seq intensities ±1kb surrounding SAR-seq peak summits for i^{3}Neurons expressing non-targeting sgRNA (sgControl, n=2) or sgRNA targeting PNKP (sgPNKP, n=2). Aggregate plots of SAR-seq intensity are shown in the top panel. **d)** Heatmap of SAR-seq ±1kb surrounding the transcription start site (TSS) in i^{3}Neurons, ordered by SAR-seq intensity. i^{3}Neurons expressing CRISPRi non-targeting sgRNAs (sgControl) or targeting PNKP (sgPNKP) were either non-treated (NT, n=2) or treated with 25 µM camptothecin (CPT, n=2) during incubation with EdU. Aggregate plots of SAR-seq intensity are shown in the top panel. **e)** Scatter plots showing correlations of intensities (RPKM) between SSBs (ddN S1-END-seq) and 5fC or 5hmC, respectively at ±1kb surrounding SAR-seq peak summits for i^{3}Neurons. Spearman correlation coefficients and p values are indicated. **f)** Scatter plots showing correlations of intensities (RPKM) between SAR-seq expressing sgPOLβ and ddN S1-END-seq, 5fC, or 5hmC at ±1kb surrounding SAR-seq peak summits for i^{3}Neurons. Spearman correlation coefficients and p values are indicated.

nature research

Corresponding author(s): André Nussenzweig

Last updated by author(s): Feb 20, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for collection of genomic sequencing data. The data was demultiplexed using blc2fastq v2.20 software from Illumina.
Data analysis	Data analysis was done publicly available software packages: bowtie v1.1.2, bowtie v2.4.1, bedtools v2.29.2, samtools v1.11, ucsc tools v407, macs v1.4.3, R v3.6.2, STAR v2.7.6a, cufflinks v2.2.1, SICER 2-1.0.2, MEME v5.3.0., deeptools v3.4.2 Genome alignment files were viewed using UCSC Genome Browser. HiC was analyzed by Juicer v1.5.6 and the interaction matrices were visualized using Juicebox Software v1.11.08. Venn diagrams were plotted by VennDiagram package in R v3.6.2. Flow cytometry data analysis was done with FlowJo v10.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw and/or processed data are available for Figures 1a,b,d,2,3,4a,b and Extended Data Figures1b-d,2a,b,d-f,3,4, 5, 6, 7b,c, 8a,b, 8d-g, 9a,c-h, 10b,d,e,h,i, 11a,c-f. Due to the size of our dataset (~110 samples, 532Gb) we were unable to submit out data to GEO directly in the timespan between being accepted in principle, and final revision. We are working with GEO to deposit our data and link them into GEO accessions.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We have tested different cell amounts for SAR-seq and found 15 million neurons was sufficient to give a good SAR-seq signal.
Data exclusions	No data were excluded from analyses.
Replication	All SAR-seq experiments have at least 2 replicates. 2 replicates for H3K4me1, XRCC1 ChIP-seq and 5hmC and 5fC SEAL. No replicates for H3K27ac, MLL4, PAR, H3K9me3, H3K4me3, H3K27me3, CTCF, H3K36me3, H3K4me1 (rat), and XRCC1 (rat) ChIP-seqs or S1-END-seq. Indicated histone mark ChIP-seqs and S1-END-seq done once to identify colocalization with SAR-seq peaks. H3K4me1 and XRCC1 ChIP-seqs only done once in rat to show that colocalization of these proteins with SAR-seq peaks is same in rat and human neurons.
Randomization	Randomization was not necessary. Cells were allocated based on genotypes confirmed by qRT-PCR/Western blot result or treatments.
Blinding	Blinding was not necessary. Knowledge of genotypes and treatments were required to suggest the mechanism and differences in SAR-seq signal were highly telling.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

anti-53BP1 (1:1000, Novus Biologicals #NB100-305, Lot# F-3)
phospho-γH2AX (1:5000, Millipore #JBW301, Lot# 3108494)
H3K4me1 (5 µg, Abcam #8895, Lot unavailable)
MLL4 (antibody courtesy of Kai Ge)
H3K27ac (5 µg, Abcam #4729, Lot# GR3357415-1)
RNA Polymerase II (8 µg, Abcam #26721, Lot# GR3305785-2)
anti-pan-PAR(5 µg, Millipore MABE1016, Lot# 3500937)
XRCC1 (2.6 µg, Novus NBP1- 87154, Lot# A117378)

anti-DNA polymerase β (1:1000, Millipore #ABE1408, Lot# 3389920)
anti-Tubulin (1:10,000, Sigma-Aldrich #T5168, Lot# unavailable)
anti-Tubulin β3 (TUBB3) (1:5000, Biolegend #801201, Lot# B264428)
H3K4me3 (5 µg, Sigma #07-473, Lot# unavailable)
H3K27me3 (5 µg, Millipore #CS200603, Lot# DAM1563757)
H3K9me3 (10 µg, Active Motif #39765, Lot# 18513004)
CTCF (6 µL, Millipore #07-729, Lot# unavailable)
H3K36me3 (5 µg, Abcam #ab9050-25, Lot# 453638)
anti-poly-PAR (1:500, Millipore MABE1031, Lot# 3384383)
DNA polymerase β (Millipore ABE1408, Lot# 3389920)
Rhodamine Red-X AffiniPure Donkey anti-Mouse IgG (H + L) (1:1000, Jackson ImmunoResearch Labs #715-295-151, Lot# unavailable)
Alexa Fluor 488 Streptavidin (1:1000, Jackson ImmunoResearch Labs #016-540-084, Lot# unavailable)
Alexa Fluor 488 goat anti-rabbit (1:1000, Invitrogen #A11034, Lot# 1812166)
Alexa Fluor 555 goat anti-mouse (1:2000, Invitrogen #A21422, Lot# 1990314)
IRDye 800 CW goat anti-rabbit (1:15,000, Li-Cor #926-32211, Lot# D00804-07)
IRDye 680 RD goat anti-mouse (1:15,000, Li-Cor #926-68070, Lot# D00804-13)

Validation

53BP1 (Novus Biologicals #NB100-305) - validated for IF by manufacturer
phospho γH2AX (Millipore #JBW301) - validated for IF by manufacturer
H3K4me1 (Abcam #8895) - validated for ChIP by manufacturer
MLL4 - validated by Kai Ge lab (PMID: 19583951)
H3K27ac (Abcam #4729) - validated for ChIP by manufacturer
RNA Polymerase II (Abcam #26721) - validated for ChIP by manufacturer
XRCC1 (Novus NBP1-87154) - validated to bind specifically to XRCC1 in by manufacturer
DNA polymerase β (Millipore ABE1408) - validated for Western blot by manufacturer
Tubulin (Sigma-Aldrich #T5168) - validated for Western blot by manufacturer
Tubulin β3 (TUBB3) (Biolegend #801201) - validated for IF by manufacturer
H3K4me3 (Sigma #07-473) - validated for ChIP by manufacturer
H3K27me3 (Millipore #CS200603) - validated for ChIP by manufacturer
H3K9me3 (Active Motif #39765) - validated for ChIP by manufacturer
CTCF (Millipore #07-729) - validated for ChIP by manufacturer
H3K36me3 (Abcam #9050) - validated for ChIP by manufacturer
anti-poly-PAR (Millipore MABE1031) - validated for IF (PMID: 32636369)
anti-pan-PAR (Millipore MABE1016) - validated for ChIP (PMID: 29053245)
Rhodamine Red-X AffiniPure Donkey anti-Mouse IgG (H + L) (1:1000, Jackson ImmunoResearch Labs #715-295-151) - validated for IF by manufacturer
Alexa Fluor 488 Streptavidin (1:1000, Jackson ImmunoResearch Labs #016-540-084) - validated for IF by manufacturer
Alexa Fluor 488 goat anti-rabbit (1:1000, Invitrogen #A11034) - validated for IF by manufacturer
Alexa Fluor 555 goat anti-mouse (1:2000, Invitrogen #A21422) - validated for IF by manufacturer
IRDye 800 CW goat anti-rabbit (1:15,000, Li-Cor #926-32211) - validated for IF by manufacturer
IRDye 680 RD goat anti-mouse (1:15,000, Li-Cor #926-68070) - validated for IF by manufacturer

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

All induced pluripotent stem cells (iPSCs) experiments used the WTC11 line, which was derived from a healthy male participant and obtained from the Coriell cell repository. We followed the policies of the NIH Intramural program for the registration and use of this iPSC line. iPSC culture was performed as described previously.

MCF10A cells are from ATCC.

Abelson-transformed murine pre-B cells was generated as described previously (Bredemeyer et al. 2006)

Authentication

The WTC11 iPSC line was not authenticated by genome sequencing, but was validated to have a normal male karyotype, as expected.

MCF10A cell line was not authenticated.

Lig4-/- v-abl lines were verified for ligase4 deficiency by PCR.

Mycoplasma contamination

iPSC lines used in this study were confirmed to be mycoplasma free based on the Lonza "MycoAlert" mycoplasma testing kit.

MCF10A cells were not tested for mycoplasma.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used in this study

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

10-11 week old wild-type, albino pregnant female rats were purchased from Envigo and delivered to our facility on day 17 of gestation. They were housed under a 12 h light-dark cycle for 24 h with access to food and water ad libitum. The following day, the

	animals were sacrificed by carbon dioxide inhalation followed by decapitation prior to d18 embryo extraction and preparation of neurons, following NIH guidelines.
Wild animals	Our study did not involve wild animals
Field-collected samples	10-11 week old pregnant albino rats were delivered to our facility on day 17 of gestation. They were housed under a 12 h light-dark cycle for 24 h with access to food and water ad libitum. The following day, the animals were sacrificed by carbon dioxide inhalation followed by decapitation prior to d18 embryo extraction and preparation of neurons. No field collected samples were used in the study.
Ethics oversight	All animal procedures were conducted following the NIH Guide for the Care and Use of Laboratory Animals, under the Animal Study Proposal #19-011 approved by the NICHD Animal Care and Use Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	Raw fastq files, aligned bigWig files, list of peaks file
Genome browser session (e.g. UCSC)	The bigwig files deposited in GEO database can be directly loaded into IGV or UCSC genome browser for visualization

Methodology

Replicates	2 replicates for H3K4me1, XRCC1 ChIP-seq and no replicates for H3K27ac, MLL4, PAR, H3K9me3, H3K4me3, H3K27me3, CTCF, H3K36me3, H3K4me1 (rat), and XRCC1 (rat) ChIP-seqs.
Sequencing depth	single-end sequencing (75 bp) at 20-30 million reads per sample for ChIP-seq
Antibodies	H3K4me1 (5 µg, Abcam #8895) MLL4 (antibody courtesy of Kai Ge) H3K27ac (5 µg, Abcam #4729) RNA Polymerase II (8 µg, Abcam #26721) anti-pan-PAR(5 µg, Millipore Sigma MABE1016) XRCC1 (2.6 µg, Novus NBP1- 87154) H3K4me3 (5 µg, Abcam #Ab8580) H3K27me3 (5 µg, Millipore #CS200603 H3K9me3 (10 µg, Active Motif #39765) CTCF (6 µL, Millipore #07-729) H3K36me3 (5 µg, Abcam #ab9050)
Peak calling parameters	We used MACS (v1.4.3) to call XRCC1, MLL4 and CTCF ChIP-seq peaks with --nomdel parameter. Histone modification ChIP-seq peaks were called by SICER with default parameters.
Data quality	fastqc was employed to examine the fastq quality.
Software	The data was aligned using bowtie v1.1.2, visualized on UCSC genome browser. Peak calling for XRCC1, MLL4 and CTCF ChIP-seq was done using MACS and SICER was used for histone modification ChIP-seq data. Heatmap was plot by deeptools.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Abelson-transformed murine pre-B cells were plated at 1e6 and treated with 1µM polymerase alpha inhibitor for 24 hours. MCF10A cells were plated at 1.5e6 and treated with 5µM Palbociclib for 48 hours, plus/minus doxycycline in the last 24 hours to induce Cas9D10A. For both cell lines, EdU (10uM) was added to cell culture medium for the final 30 minutes, before cells were washed with PBS and harvested by trypsinization. Cells were centrifuged at 500xg for 5 minutes, washed once with PBS, and subsequently fixated in 100uL 4% PFA for 30 minutes. After fixation cells were pelleted, washed with PBS, and permeabilized using saponin-based reagent (Click-IT EdU Alexa Fluor 488 Flow Cytometry Assay Kit (ThermoFisher)). Cells were again washed in saponin reagent and incubated in EdU-Click It mix for 30 minutes as per manufacturer's instructions Click-IT EdU Alexa Fluor 488 Flow Cytometry Assay Kit (ThermoFisher). Finally, cells were washed with PBS+2%FBS, resuspended in PBS+DAPI (500ng/mL) and left in the dark at RT for at least 15 minutes before running samples on flow cytometer.
Instrument	BD FACSCanto II
Software	FlowJo v10
Cell population abundance	20,000 cells per sample
Gating strategy	Cell debris and dead cells were excluded based on forward scatter area (FCS-A) and side scatter area (SSC-A), cell doublets were excluded based on forward scatter area (FCS-A) and forward scatter height (FCS-H). Cell cycle analysis was performed based on EdU-AF488 or EdU-AF647 and DAPI (DNA content) signal; EdU-488 or EdU-647 positive cells were identified as S-phase. EdU-AF488 or EdU-AF647 negative cells were identified as either G1 (low DAPI signal) or G2/M (high DAPI signal).

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.